

## PROTEASE

### FIELD OF THE INVENTION

The present invention relates to IgM specific proteases. Specifically, proteases according to  
5 the present invention are able to cleave IgM, including human IgM, with high specificity. The  
invention further relates to methods and uses of said proteases, particularly for the analysis  
and/or modification of IgM.

### BACKGROUND OF THE INVENTION

Pathogenic bacteria have evolved numerous mechanisms to evade host immune systems,  
10 which allow them to successfully colonize a desired host environment.

One such mechanism is the ability of some bacteria to modulate the functionality of host  
antibodies in order to render them less active or inactive. For example, some known  
bacterially-encoded virulence factors have IgG specific glycolytic or proteolytic activity  
15 (Sjögren J., et al., *Glycobiology* 2020, 30(4):254-267; Brezski RJ, Jordan RE. *mAbs* 2010,  
2(3):212-220), and others have IgA specific proteolytic activity (Kornfeld SJ, Plaut AG. *Rev*  
*Infect Dis* 1981, 3(3):521-534).

Of the known virulence factors that modulate the host immune system through enzymatic  
20 modification of host proteins, most typically derive from zoonotic or species-restricted  
pathogens. Indeed, IgG and IgM specific proteases have been isolated from pathogenic  
species within the *Streptococcus* genus (Von Pawel-Rammingen U., et al., *EMBO* 2002,  
21(7):1607-1615; Lannergård J & Guss B., IdeE, *FEMS Microbiol Lett* 2006, 262(2):230-5;  
Spoerry C., et al., *J Biol Chem* 2016, 291(15):7915-25; Seele J., et al., *J Bacteriol* 2013,  
25 195(5)).

The activity and importance of IgG is well known and widely exploited in biotechnology and  
medicine. IgM, on the other hand, is often less well considered but is nonetheless a critical  
component of the immune system. In response to microbial infection, for example, IgM plays  
30 an important role in maintaining an intact primary barrier, wherein IgM can neutralize  
microbes by inhibiting the binding between oral and respiratory microorganism and their  
cognate cellular receptors. Further, IgM plays a key role in the opsonization of pathogens  
and activation of complement.

IgM has further uses in biotechnology and medicine, with applications that extend beyond responding to microbial infection. Such uses are becoming increasingly common and attractive for therapeutic exploitation (Samsudin F., et al., *Chem Sci* 2020, 11(10):2843-2854; Zhang J., et al., *mAbs* 2022, 14(1):2031483).

5

Despite the importance of IgM in both the endogenous functioning of the immune system and in therapeutic and biotechnological applications, there are a lack of existing tools for the detection, analysis, and modification of IgM that are efficient, widely applicable, and target specific.

10

An IgM specific protease has been identified in the zoonotic porcine pathogen *Streptococcus suis* (Seele J., et al., *J Bacteriol* 2013, 195(5)). The activity of this protease is species restricted to substrates including porcine IgM, however it is incapable of cleaving human IgM.

15

As such, it is evident that there are a paucity of biological tools suitable for the study and manipulation of IgM. In particular, there is an unmet need for biological tools that specifically target human IgM.

## 20 SUMMARY OF THE INVENTION

The present invention relates to the surprising finding that polypeptides isolated from the *Lachnoanaerobaculum* genus of commensal human bacteria are able to cleave IgM with high specificity. As such, the polypeptides of the invention allow for the specific and effective targeting of IgM, including human IgM. The polypeptides of the invention have utility in  
25 numerous applications as biological tools for the study and manipulation of IgM.

In a first aspect, there is provided a polypeptide having human IgM specific endoprotease activity.

30 In one aspect, there is provided a polypeptide having endoprotease activity comprising:

- a) the amino acid sequence of SEQ ID NO: 1;
- b) a variant thereof having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 1; or
- c) a fragment of either a) or b),

preferably wherein said endoprotease activity is human IgM specific endoprotease activity.

In another aspect, there is provided a polypeptide having endoprotease activity comprising:

- a) the amino acid sequence of SEQ ID NO: 8;
- 5        b) a variant thereof having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 8; or
- c) a fragment of either a) or b),

preferably wherein said endoprotease activity is human IgM specific endoprotease activity.

- 10       In one aspect, there is provided a composition or combination comprising a first polypeptide according and a second polypeptide, wherein the first polypeptide is a polypeptide having endoprotease activity according to the invention, and the second polypeptide is a polypeptide having IgG specific endoprotease activity.

- 15       In one aspect, there is provided a polynucleotide encoding a polypeptide according to the invention.

In one aspect, there is provided a vector encoding or comprising a polynucleotide according to the invention.

20

In one embodiment, there is provided an expression vector encoding or comprising a polynucleotide encoding a polypeptide according to the invention, wherein the expression vector is for the production of a polypeptide according to the invention.

- 25       In one aspect, there is provided a kit comprising:
  - a) the polypeptide according to the invention
  - b) the polynucleotide according to the invention
  - c) the vector according to the invention; and/or
  - d) the composition or combination according to the invention, wherein the first
  - 30           polypeptide and the second polypeptide of said combination are provided separately; and/or
  - e) the composition or combination according to the invention, wherein the first
  - polypeptide and the second polypeptide of said combination are provided together.

35

In one aspect, there is provided a method of hydrolysing IgM, the method comprising contacting a sample comprising IgM with the polypeptide, the composition, or the combination according to the invention.

- 5 In one aspect, there is provided a method of detecting the presence of IgM in a sample, the method comprising:
- a) contacting a sample with the polypeptide according the invention; and
  - b) identifying and/or isolating the IgM specific cleavage products,

wherein the presence of IgM specific cleavage products is indicative of the presence of IgM  
10 in the sample.

In one aspect, there is provided the use of the polypeptide, the combination, or the kit according to the invention for the degradation, removal, or identification of IgM in a sample.

- 15 In one aspect, there is provided a polypeptide, composition, combination, pharmaceutical composition, or kit for use in therapy.

## BREIF DESCRIPTION OF THE FIGURES

### Figure 1: SDS-PAGE analysis of LU-M1

- 20 SDS-PAGE analysis of a recombinantly expressed modified version of a polypeptide derived from *L. umeaense* termed "LU-M1" demonstrates the polypeptide is produced at high yield and high purity. Apparent molecular mass is indicated by reference to the molecular weight ladder (1<sup>st</sup> and 9<sup>th</sup> lanes); the molecular weight of ladder components are shown. The "elution" lane demonstrates the ability to obtain a highly pure and homogenous sample of  
25 LU-M1 after a single round of affinity purification using the His-tag.

### Figure 2: *Lachnoanaerobaculum* derived polypeptides represent the first known human IgM specific endoproteases

- The activity of the different constructs were assessed using the full-length constructs from *L.*  
30 *umeaense* (LU-FL) and *L. gingivalis* (LG-FL) towards IgG from human and mouse (A). LG-FL was further evaluated towards subclasses of human IgG (B). Both LG-FL and LU-FL were mixed with human IgA and IgM (C). Activity of the LU-M1 construct was assessed by mixing it with all human IgG subclasses as well as polyclonal IgG, IgA, and IgM (D). General

proteolytic activity of LU-M1 was investigated using a fluorescent casein substrate (E), and species specificity using IgM purified from mouse, rabbit, rat, and monkey (F). SpeB (general protease) and IdeS (IgG-specific protease) were used as controls. Figures 2A-D and 2F depict SDS-PAGE gels, lanes are labelled accordingly and apparent molecular mass is indicated by reference to the molecular weight ladder.

**Figure 3: *Lachnoanaerobaculum* derived polypeptides catalyze site specific cleavage of IgM**

Liquid chromatography mass-spectrometry (LC-MS) analysis of PNGaseF treated, reduced, and denatured myeloma IgM. The total ion current chromatogram (TIC; top panel) and the deconvoluted compounds of interest (lower panels) is shown for samples that were not further treated (Figure 3A) or further treated with LU-M1 (Figure 3B).

**Figure 4: Size exclusion analysis of IgM cleavage products**

Size exclusion chromatography of the hydrolyzed IgM sample was performed alongside size standards consisting of Thyroglobulin (1), Apoferritin (2), and Human IgG (3). Intact IgM, as well as the hydrolysis products (pentameric CH3-CH4; CH1-CH1-CH2) are indicated on the chromatogram.

**Figure 5: Assessment of catalytic activity**

The activity of LU-M1 was assessed under a variety of conditions, which include a range of pH (A), NaCl concentration (B), the presence of divalent cations (C), temperature (D), and in the presence of protease inhibitors (E). Figures 5A-E depict SDS-PAGE gels, lanes are labelled accordingly and apparent molecular mass is indicated by reference to the molecular weight ladder.

**Figure 6: LU-M1 is a highly efficient enzyme**

The rate of LU-M1 activity on IgM was assessed. IgM was mixed with the LU-M1 at varying enzyme to substrate ratios (1:20 to 1:100; w:w), whilst total reaction volume remained constant. Reactions were incubated at 37°C for between 5 and 60 minutes. Figure 6A depicts SDS-PAGE gels, lanes are labelled accordingly and apparent molecular mass is

indicated by reference to the molecular weight ladder. Figures 6B and 6C depict densitometric data obtained from corresponding SDS-PAGE analysis.

**Figure 7: LU-M1 is active in serum**

- 5 Human serum was incubated with IdeS and LU-M1 and incubated for 30 min at 37°C before being analyzed by SDS-PAGE (top panels), and western blotting against IgG (A) and IgM (B) (bottom panels).

**Figure 8: LU-M1 can be lyophilized and immobilized and retain activity**

- 10 Lyophilized LU-M1 was reconstituted with water to identical concentrations, and the activity assessed against IgM (Figure 8A). The numbers above the SDS-PAGE indicates the lyophilization conditions of the sample prior to reconstitution. (B) Image of lyophilization pellets. The activity of resin immobilized LU-M1 was assessed against IgM (Figure 8C). The numbers above the SDS-PAGE indicates the concentration of LU-M1 immobilized. Figures  
15 8A and 8C depict SDS-PAGE gels, lanes are labelled accordingly and apparent molecular mass is indicated by reference to the molecular weight ladder.

**Figure 9: LU-M1 is not readily recognized by human IVIG**

- The presence of LU-M1 specific human IgG was assessed using western blotting. 0.5 µg  
20 each of IdeS, Xork, and LU-M1 were separated by SDS-PAGE and either stained (left panel) or transferred to a nitrocellulose membrane for western blotting (right panel). The membrane was blocked, incubated with human IVIG, and an alkaline-phosphatase conjugated anti-human IgG Fc antibody was used for detection. Western blot data indicate that little to no anti-LU-M1 antibodies exist in IVIG.

25

**DETAILED DESCRIPTION OF THE INVENTION**

- It is to be understood that different applications of the disclosed products and methods may be tailored to the specific needs in the art. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only,  
30 and is not intended to be limiting.

All publications, patents and patent applications cited herein are hereby incorporated by reference in their entirety.

5 Unless specified otherwise, terms will be understood as having a meaning that is common in the art as would be attributed to them by the person skilled in the art.

The terms "comprising", "comprises" and "comprised of" as used herein are synonymous with "including", "includes" or "containing", "contains", and are inclusive or open-ended and do not exclude additional, non-recited members, elements or method steps. The terms  
10 "comprising", "comprises" and "comprised of" also include the term "consisting of". As such, the term "including" is used to mean "including but not limited to". "Including" and "including but not limited to" may be used interchangeably. Conversely, the term "consisting of" is intended to be exclusive or closed and does not include additional, non-recited members, elements or method steps.

15 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. Thus, for example, reference to "a polypeptide" includes "polypeptides", and the like.

## 20 **IgM**

IgM is an immunoglobulin (Ig) isotype that is typically characterised by its large size and pentameric structure. However, IgM may also exist in other structural forms such as a secreted hexamer or cell-surface displayed monomer that forms part of the B-cell receptor.

25 The pentameric structure of IgM is such that it is made up essentially of five structures analogous to IgG. IgM comprises 10 potential antigen-binding sites (or paratopes), which facilitate high avidity target binding. This high avidity binding can allow target binding even when monovalent binding affinities are low.

30 IgM is produced early during infection following antigen exposure and plays a key role in stimulating other effector functions of the immune system, such as activating complement and Fc receptor mediated activities.

IgM has been demonstrated to have favorable therapeutic characteristics, even when  
35 compared to IgG, and has been described as a promising candidate for monoclonal antibody

therapies for several conditions (Samsudin F., et al., *Chem Sci* 2020, 11(10):2843-2854; Zhang J., et al., *mAbs* 2022, 14(1):2031483).

### **IgM specific endoproteases**

- 5 The invention concerns polypeptides having IgM specific endoprotease activity.

It will be understood that, as used herein, the term "polypeptide having IgM specific endoprotease activity" may refer to any protein, polypeptide, or fragment thereof, that is able to catalyze the cleavage of IgM. The terms endoprotease and protease may be used  
10 interchangeably herein.

Determination of whether cleavage of IgM has occurred may be readily performed by the person skilled in the art, for example using common techniques in the art of protein biochemistry, such techniques including SDS-PAGE analysis, western blotting,  
15 chromatography (e.g., size-exclusion chromatography), and mass spectrometry. Such techniques are exemplified herein (see Examples) but are in no way intended to be limiting.

Further, it will be understood that IgM specific endoprotease activity may refer to endoprotease activity that is target (IgM) specific. Said activity may not be considered to  
20 have a broad spectrum of target proteins, for example as trypsin does. Nevertheless, IgM specific activity encompasses activity against different IgM molecules, for example, IgM derived from different species or modified IgM molecules.

Modified IgM molecules may include, for example, non-naturally occurring (e.g., synthetic)  
25 IgM molecules; fusion proteins comprising IgM molecules or parts/sequences thereof; and fragments of IgM molecules.

The skilled person will appreciate that the polypeptide may have residual or lower level activity for other targets, such as IgG.

30 In one embodiment, the polypeptide may have IgM specific and IgG specific protease activity.

In some embodiments, the polypeptide may have low or negligible levels of detectable IgG  
35 protease activity.

In some embodiments, the polypeptide may have low levels of detectable IgG activity, wherein the level of IgG activity is less than the level of IgM activity.

5 In some embodiments, the polypeptide may have low levels of detectable IgG activity, wherein the level of IgG activity is less than 70% of the level of IgM activity, such as less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, or less than 5% of the level of IgM activity.

10 In some embodiments, the IgG activity is IgG<sub>2</sub> specific activity.

In a preferred embodiment, the polypeptide has no detectable activity against polypeptides other than IgM.

15 In another preferred embodiment, the polypeptide has no detectable activity against any immunoglobulin isotype other than IgM.

In another preferred embodiment, the polypeptide has no detectable activity against IgG.

20 In a preferred embodiment of the invention, the IgM specific endoprotease activity is human IgM specific endoprotease activity.

25 It will be understood that the term "human IgM specificity" does not imply that the polypeptide is restricted to human IgM, rather, the endoprotease is specific for IgM and human IgM is a recognised substrate, but not necessarily the only substrate. For example, IgM from non-human primates may also be a substrate.

In one embodiment the polypeptide activity against IgM derived from monkeys of the genus *Macaca*, such as *Macaca mulatta* and/or *Macaca fascicularis*.

30 In one embodiment the polypeptide activity against IgM derived from monkeys of the genus *Papio*.

#### Lachnoanaerobaculum

35 *Lachnoanaerobaculum* is a genus of human commensal bacteria that are found in the human mouth and intestines. *Lachnoanaerobaculum orale* and *Lachnoanaerobaculum umeaense* were first described in 2012, following isolation of from the oral cavity of a healthy

individual and from the intestine of a child with coeliac disease, respectively (Hedberg M.E., et al., *J Syst Evol Microbiol* 2012, 62(Pt11):2685-2690).

Despite two very recent studies that implicate *Lachnoanaerobaculum* species, *L. orale* and  
5 *L. gingivalis*, in causing bacteremia in immune-compromised individuals (Ida Y., et al.,  
*Anaerobe* 2022, 73:102506; Okada N., et al., *Anaerobe* 2022, 76:102610),  
*Lachnoanaerobaculum* species have not been described as the causative agent of any  
pathologies.

10 Until the work of the present inventors, as described herein, no isolate from the  
*Lachnoanaerobaculum* genus has been demonstrated to have proteolytic activity, nor has  
any specific immunoglobulin-modifying activity been identified or investigated.

Surprisingly, the present inventors have identified IgM specific endoproteases encoded by  
15 bacteria of the *Lachnoanaerobaculum* genus.

In one embodiment, the invention provides a polypeptide having IgM specific endoprotease  
activity, wherein the polypeptide is derived from a bacterium of the *Lachnoanaerobaculum*  
genus.

20 In one embodiment, the invention provides a polypeptide having IgM specific endoprotease  
activity, wherein the polypeptide is encoded by a polynucleotide sequence that has at least  
80% sequence identity, such as at least 85%, at least 90%, at least 95%, at least 98%, or at  
least 99% sequence identity, to a polynucleotide sequence encoded by  
25 *Lachnoanaerobaculum umeaense*.

In one embodiment, the invention provides a polypeptide having IgM specific endoprotease  
activity, wherein the polypeptide is encoded by a polynucleotide sequence that has at least  
80% sequence identity, such as at least 85%, at least 90%, at least 95%, at least 98%, or at  
30 least 99% sequence identity, to a polynucleotide sequence encoded by  
*Lachnoanaerobaculum gingivalis*.

In one embodiment, the invention provides a polypeptide having IgM specific endoprotease  
activity, wherein the polypeptide has at least 80% sequence identity, such as at least 85%,  
35 at least 90%, at least 95%, at least 98%, or at least 99% sequence identity, to an amino acid  
sequence encoded by *Lachnoanaerobaculum umeaense*.

- In one embodiment, the invention provides a polypeptide having IgM specific endoprotease activity, wherein the polypeptide has at least 80% sequence identity, such as at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% sequence identity, to an amino acid sequence encoded by *Lachnoanaerobaculum gingivalis*.
- 5 In one embodiment, the invention provides a polypeptide having IgM specific endoprotease activity, wherein the polypeptide is derived from *Lachnoanaerobaculum umeaense*.
- In one embodiment, the invention provides a polypeptide having IgM specific endoprotease activity, wherein the polypeptide is derived from *Lachnoanaerobaculum gingivalis*.
- 10

- Polypeptides of the invention may be fragments or modified versions of naturally occurring polypeptides encoded by bacteria of the *Lachnoanaerobaculum* genus.
- 15 Engineered IgM specific endoproteases
- Following identification of candidate protease genes in *Lachnoanaerobaculum* species, the present inventors engineered a number of polypeptides.
- A polypeptide according to the present invention may comprise or consist of a sequence derived from *Lachnoanaerobaculum umeaense*.
- 20
- Exemplary *Lachnoanaerobaculum umeaense* polypeptide sequences are set out in table 1 herein.

25 Table 1: Exemplary *Lachnoanaerobaculum umeaense* sequences

| Sequence identifier | Polypeptide sequence                                                                                                                                                                                                                                                                                                                                                    |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SEQ ID NO: 1        | NIVESDQDKANRIKSSLTTVKNDNVVQYTNEDGEIRTIWAKGINGPTM<br>GEGGDFHKEITNAGSDTYIDYKTSFTTGDGWYDVNKTRAGGNVDVDK<br>NLCFGAVSSNMLHWWFEQNI EYIDRYIVKNGDIIRANRKISDLKYSFKS<br>QQNSAIFELFKVLFGNNEKGFYTDLLMDLFINGYTPKTGGGTNIERDDL<br>TPDSRGGFFYDVFKGEKLTDRTYGGSYEALGDKLKEVLGEAGLIGISH<br>KAFKNSNHIVTLWGA EYDLNGKLI AVYISDSDDQNETEDERDLGMKRY<br>EVRNVGGVAKLSTNQTNKSSGSAIGY LHILYLGTTQWENYFK |
| SEQ ID NO: 2        | MKKFNGKSVFLGLAAATI ISSNAYAGIEGWKFEGNNWKYYKEDKALK A<br>WQQLNNKWYFFNEDGSLKTGWHL DATNNWYFLDSSKNSNEGMLLSG<br>WQWIDGYCYFFENIDKSTLGRMYVNTVVNGYRIDATGRWTNEKGVEY<br>YEAGKGIITVNTLQGSLNTGKKSSGGGSSGKGGGLGKSGESGKGGSS                                                                                                                                                             |

|              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|              | IGGIGGKTGINQGDNDTSNQSLIVEENSTVNNADQTKSVEKTKGDQK<br>SEQNNSNEVNTTEGGQNKSGMTNNTSEYTYNKPTVNKENEDEKTNTT<br>QDNKVSdstSNDLSKTKDTQTNTNDSDIKSATASNAEIQKEEEHKNT<br>TTNSSEKNKNTQQNKSEENIVESDQDKANRIKSSLTTVKNDNVVQYT<br>NEDGEIRTIIWAKGINGPTMGEggDFHKEITNAGSDTYIDYKTSFTTGD<br>GWYDVNKTRAGGNVDVDKNLCFGAVSSNMLHWWFEQNIEYIDRYIVK<br>NGDIIRANRKISDLKYSFKSQQNSAIFELFKVLFGNNEKGFYTDLLMDLF<br>INGYTPKTGGGTNIERDDLTPDSRGGFFYDVFKEKLTDRTYGGSYEA<br>LGDKLKEVLGEAGLIGISHKAFKNSNHIVTLWGAEYDLNGKLIAYVISDS<br>DDQNETEDERDLGMKRYEVRNVGGVAKLSTNQTNKSSGSAIGYLHILY<br>LGTTQWENYFK                                                                                                                                                                                                        |
| SEQ ID NO: 3 | MGIEGWKFEGNNWKYYKEDKALKAWQQLNNKWYFFNEDGSLKTGW<br>HLDATNNWYFLDSSKNSNEGMLLSGWQWIDGYCYFFENIDKSTLGRM<br>YVNTVVNGYRIDATGRWTNEKGVYEEAGKGIITVNTLQGSINTGKKS<br>SGGGSSGKGGGLGKSGESGKGGSSIGGIGGKTGINQGDNDTSNQSL<br>LIVEENSTVNNADQTKSVEKTKGDQKSEQNNSNEVNTTEGGQNKSGM<br>TNNTSEYTYNKPTVNKENEDEKTNTTQDNKVSdstSNDLSKTKDTQT<br>NDSDIKSATASNAEIQKEEEHKNTETTNSSEKNKNTQQNKSEENIVE<br>SDQDKANRIKSSLTTVKNDNVVQYTNEEDGEIRTIIWAKGINGPTMGEgg<br>DFHKEITNAGSDTYIDYKTSFTTGDGWYDVNKTRAGGNVDVDKNLCF<br>GAVSSNMLHWWFEQNIEYIDRYIVKNGDIIRANRKISDLKYSFKSQQNS<br>AIFELFKVLFGNNEKGFYTDLLMDLFINGYTPKTGGGTNIERDDLTPDS<br>RGGFFYDVFKEKLTDRTYGGSYEA LGDKLKEVLGEAGLIGISHKAFK<br>NSNHIVTLWGAEYDLNGKLIAYVISDSDDQNETEDERDLGMKRYEVRN<br>VGGVAKLSTNQTNKSSGSAIGYLHILYLGTTQWENYFKHHHHHH                 |
| SEQ ID NO: 4 | MHHHHHHHHHHGIEGWKFEGNNWKYYKEDKALKAWQQLNNKWYFF<br>NEDGSLKTGWHLDATNNWYFLDSSKNSNEGMLLSGWQWIDGYCYFF<br>ENIDKSTLGRMYVNTVVNGYRIDATGRWTNEKGVYEEAGKGIITVNTL<br>QGSINTGKKS SGGGSSGKGGGLGKSGESGKGGSSIGGIGGKTGINQ<br>GDNDTSNQSLIVEENSTVNNADQTKSVEKTKGDQKSEQNNSNEVNT<br>TEGGQNKSGMTNNTSEYTYNKPTVNKENEDEKTNTTQDNKVSdstSN<br>DLSKTKDTQTNTNDSDIKSATASNAEIQKEEEHKNTETTNSSEKNKNT<br>QQNKSEENIVESDQDKANRIKSSLTTVKNDNVVQYTNEEDGEIRTIIWAK<br>GINGPTMGEggDFHKEITNAGSDTYIDYKTSFTTGDGWYDVNKTRAG<br>GNVDVDKNLCFGAVSSNMLHWWFEQNIEYIDRYIVKNGDIIRANRKISD<br>LKYSFKSQQNSAIFELFKVLFGNNEKGFYTDLLMDLFINGYTPKTGGGT<br>NIERDDLTPDSRGGFFYDVFKEKLTDRTYGGSYEA LGDKLKEVLGEA<br>GLIGISHKAFKNSNHIVTLWGAEYDLNGKLIAYVISDSDDQNETEDERDL<br>GMKRYEVRNVGGVAKLSTNQTNKSSGSAIGYLHILYLGTTQWENYFK<br>HHHHHH |
| SEQ ID NO: 5 | MQNKSGMTNNTSEYTYNKPTVNKENEDEKTNTTQDNKVSdstSNDLS<br>KTKDTQTNTNDSDIKSATASNAEIQKEEEHKNTETTNSSEKNKNTQQ<br>NKSEENIVESDQDKANRIKSSLTTVKNDNVVQYTNEEDGEIRTIIWAKGI<br>NPTMGEggDFHKEITNAGSDTYIDYKTSFTTGDGWYDVNKTRAGGNV<br>VDVDKNLCFGAVSSNMLHWWFEQNIEYIDRYIVKNGDIIRANRKISDLKY<br>SFKSQQNSAIFELFKVLFGNNEKGFYTDLLMDLFINGYTPKTGGGTNIE<br>RDDLTTPDSRGGFFYDVFKEKLTDRTYGGSYEA LGDKLKEVLGEAGLI<br>GISHKAFKNSNHIVTLWGAEYDLNGKLIAYVISDSDDQNETEDERDLGM<br>KRYEVRNVGGVAKLSTNQTNKSSGSAIGYLHILYLGTTQWENYFKHHH<br>HHH                                                                                                                                                                                                                                                               |
| SEQ ID NO: 6 | MSDSTSNDLSKTKDTQTNTNDSDIKSATASNAEIQKEEEHKNTETTNS<br>SEKNKNTQQNKSEENIVESDQDKANRIKSSLTTVKNDNVVQYTNEDE<br>GIRTIIWAKGINGPTMGEggDFHKEITNAGSDTYIDYKTSFTTGDGWYD<br>VNKTRAGGNVDVDKNLCFGAVSSNMLHWWFEQNIEYIDRYIVKNGDI<br>RANRKISDLKYSFKSQQNSAIFELFKVLFGNNEKGFYTDLLMDLFINGY                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |

|              |                                                                                                                                                                                                                                                                                                                                                                         |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|              | TPKTGGGTNIERDDLTPDSRGGFFYDVFKEKLTDRTYGGSYEALGD<br>KLKEVLGEAGLIGISHKAFKNSNHIVTLWGAEYDLNGKLIAYISDSDQ<br>NETEDERDLGMKRYEVRNVGGVAKLSTNQTNKSSGSAIGYLHILYLG<br>TQWENYFKLEHHHHHH                                                                                                                                                                                               |
| SEQ ID NO: 7 | MNIVESDQDKANRIKSSLTTVKNDNVVQYTNEDEGEIRTIWAKGINGPT<br>MGEAGDFHKEITNAGSDTYIDYKTSFTTGDGWYDVNKTRAGGNVDVD<br>KNLCFAGVSSNMLHWWFEQNIYIDRYIVKNGDIIRANRKISDLKYSFKS<br>QQNSAIFELFKVLFGNNEKGFYTDLLMDLFINGYTPKTGGGTNIERDDL<br>TPDSRGGFFYDVFKEKLTDRTYGGSYEALGDKLKEVLGEAGLIGISH<br>KAFKNSNHIVTLWGAEYDLNGKLIAYISDSDQNETEDERDLGMKRY<br>EVRNVGGVAKLSTNQTNKSSGSAIGYLHILYLGTTQWENYFKHHHHHH |

A polypeptide according to the present invention may comprise or consist of a sequence derived from *Lachnoanaerobaculum gingivalis*.

- 5 Exemplary *Lachnoanaerobaculum gingivalis* polypeptide and polynucleotide sequences are set out in table 2 herein.

Table 2: Exemplary *Lachnoanaerobaculum gingivalis* sequences

| Sequence identifier | Polypeptide sequence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SEQ ID NO: 8        | AERKESTQNRAERFKESLITSENDNVVQYTNEDEGEIRTIWVKGINGPV<br>MGEAGDFHKEITQGGKNTYVMYTPYSSGDGWYDVNKTRSGGNIDID<br>KNLCFAAVSSNMLHWWFDQNIYVDRIEKNNGDIIRANRKLSDLKTSFE<br>NPENSKIFELYKVLYGYNERGFYSDLLMDLFINGYTPNPKGGTNIESDD<br>LIPNNNGGFFYDVFKEGKLTDRTYGGNYESLSNLLKEILGNGGIVGLSH<br>KVFSKSNHIVTLWGAEYDLNGKLVAVYISDSDQDES NVGMKRFEIRN<br>VGGKAKISTNETDKSAGAEGLYLHILYQGTGQWENYFK                                                                                                                                                                                                                                                                                                                                                  |
| SEQ ID NO: 9        | MKKYKVG PVLLGLTMLMSFNAYAGVDGWKQENNSWKYYKEDKALKA<br>WQQINNNWYFFNEDGSLKTGWYMDAANNWYFLDSSKTANEGVLLSG<br>WQWIDGYCYFFEGIDNNTFGKMYSNLTVSGYKVDALGRWINESGAEY<br>YEAGKGISTGNASQSSSNTAKKTSVAGSSEKSGKSGGSSKGGSSGKS<br>GGSSKGGSSGKSGGSSFLNQGEKREENRILGESKGAENSSSKRSED<br>LVGSRQDLSSDKNNSLIADNKTGNSEKAGSIEKNTNKEYTTKSNEPV<br>KPATSSNVEIPKEEEKKDIESKESKKPEAERKESTQNRAERFKESLITSE<br>NDNVVQYTNEDEGEIRTIWVKGINGPVMGEAGDFHKEITQGGKNTYVM<br>YTPYSSGDGWYDVNKTRSGGNIDIDKNLCFAAVSSNMLHWWFDQNI<br>EYVDRIEKNNGDIIRANRKLSDLKTSFENPENSKIFELYKVLYGYNERGF<br>YSDLLMDLFINGYTPNPKGGTNIESDDLIPNNNGGFFYDVFKEGKLTDR<br>TYGGNYESLSNLLKEILGNGGIVGLSHKVFSKSNHIVTLWGAEYDLNGK<br>LVAVYISDSDQDES NVGMKRFEIRNVGGKAKISTNETDKSAGAEGLY<br>LHILYQGTGQWENYFK |
| SEQ ID NO:<br>10    | MGVDGWKQENNSWKYYKEDKALKAWQQINNNWYFFNEDGSLKTGW<br>YMDAANNWYFLDSSKTANEGVLLSGWQWIDGYCYFFEGIDNNTFGK<br>MYSNLTVSGYKVDALGRWINESGAEY YEAGKGISTGNASQSSSNTAK<br>KTSVAGSSEKSGKSGGSSKGGSSGKSGGSSKGGSSGKSGGSSFLNQ<br>GEKREENRILGESKGAENSSSKRSEDLVGSRQDLSSDKNNSLIADNK                                                                                                                                                                                                                                                                                                                                                                                                                                                         |

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                  | TGNSEKAGSIEKNTNKEYTTKSNESPVKPATSSNVEIPKEEEKKDIESK<br>ESKKPEAERKESTQNRAERFKESLITSENDNVVQYTNEDGEIRTIWVK<br>GINGPVMGEGGDFHKEITQGGKNTYVMYTPYSSGDGWYDVNKTRS<br>GGNIDIDKNLCFAAVSSNMLHWWFDQNIEYVDRIEKNGDIIIRANRKL<br>DLKTSFENPENSIFELYKVLYGYNERGFYSDLLMDLFINGYTPNPKGG<br>TNIESDDLIPNNNGGFFYDVFKGKKLTDRTYGGNYESLSNLLKEILGNG<br>GIVGLSHKVFSKSNHIVTLWGAEYDLNGKLVAVYISDSDDQDESNVGM<br>KRFEIRNVGGKAKISTNETDKSAGAELGYLHILYQGTGQWENYFKHHH<br>HHH                                                                                                                                                                                                                                                           |
| SEQ ID NO:<br>11 | MHHHHHHHHHHGVDGWKQENNSWKYYKEDKALKAWQQINNNWYFF<br>NEDGSLKTGWYMDAANNWYFLDSSKTANEGVLLSGWQWIDGYCYFF<br>EGIDNNTFGKMYSNTLVSGYKVDALGRWINESGAEYEEAGKGISTGNA<br>SQSSSNTAKKTSVAGSSEKSGKSGSSKGGSSGKSGSSKGGSSGK<br>SGSSFLNQGEKREENRILGESKGAENSSSKRSEDLVGSRQDLSSD<br>KNNSLIADNKTGNSEKAGSIEKNTNKEYTTKSNESPVKPATSSNVEIPK<br>EEEEKDIESKESKKPEAERKESTQNRAERFKESLITSENDNVVQYTN<br>DGEIRTIWVKINGPVMGEGGDFHKEITQGGKNTYVMYTPYSSGDG<br>WYDVNKTRSGGNIDIDKNLCFAAVSSNMLHWWFDQNIEYVDRIEKN<br>GDIIRANRKLSDLKTSFENPENSIFELYKVLYGYNERGFYSDLLMDLFI<br>NGYTPNPKGGTNIESDDLIPNNNGGFFYDVFKGKKLTDRTYGGNYESL<br>SNLLKEILGNGGIVGLSHKVFSKSNHIVTLWGAEYDLNGKLVAVYISDS<br>DDQDESNVGMKRFEIRNVGGKAKISTNETDKSAGAELGYLHILYQGTG<br>QWENYFK |

Exemplary polynucleotide sequences that encode polypeptides according to the invention are set out in Table 3.

5 Table 3: Exemplary *Lachnoanaerobaculum* polynucleotide sequences

| Sequence identifier | Polynucleotide sequence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Corresponding polypeptide sequence encoded thereby |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|
| SEQ ID NO:<br>12    | aatatagtagagagtgatcaggataaggccaatagaataaaatcc<br>tctttgactaccgttaaaaatgataatgtggtacagtacacgaatga<br>agatggagaaattagaacaattatttgggcaaaggcatcaatgg<br>acctactatgggagaggggtggagatttcacaaagaaataacaaa<br>tgcgggttcagatacatatagattataaaacttcgttcactacaggt<br>gatggatggtatgatgaaataagacaagagccggtggcaatgta<br>gacgtagataaaaatcttgccttggggcagttcatccaatgctac<br>attggtggttgaacaaaatattgaatatattgacagatatatagtaa<br>agaatggagatataattcgtgcaaataaggaagatctcagatctgaa<br>atattctttaagagtcagcaaaatagtgctatatcgaactcttaag<br>gtgttatttgaataatgagaagggtatttatacagacctttgatgg<br>acctgtttataaacgatatactccaaagacaggtggaggtacaaa<br>tatagaaagagatgatttaacacctgatagcagaggtggattttcta<br>tgatgtgtttaaaggagagaagctgacagaccgtactatgggtga<br>agttatgaagcttaggtgataaataaaagaagtactgggagaag<br>ccggtttgataggtatatctcataaggcttttaaaacagtaatatat | SEQ ID NO: 1                                       |

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |              |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
|                  | agtaactctatggggtgctgaatatgattggaatgaaagccttatagc<br>ggtatatattagtgattcagatgatcaaaatgaaacagaagatgaa<br>agagatctgggcatgaaacgctatgaggtaggaatgtaggagga<br>gtggcaaaactgtccacaaatcaaacgaataagtcgtcagggtca<br>gctataggatatctcatatactatacttaggaaccacacagtggga<br>aaattattttaaataa                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |              |
| SEQ ID NO:<br>13 | atgaaaaagttaacggtaagtctgttttttaggacttgcagcagca<br>actataatctcatctaatagcctatgcaggattgaaggatggaaatt<br>gagggtaataattgaaatattataaagaagataaagctttaaagg<br>cctggcagcaattaaataataaattggtatttttaatgaggatggaa<br>gtttaaaaacaggatggcatttggatgctacaaataattggtactttt<br>agacagctctaaaaattctaatagaaggtatgctttgtcaggatggc<br>agtggatagacggatactgctatttttgaacacattgataaatcta<br>ctcttggcagaatgtatgaaatactgtagtgaatggatatagaattg<br>atgtacaggtatagtgacaaatgaaaaagggtgtgaatattatga<br>agcaggcaagggaataattacagtaatactttacaaggctctttaa<br>atacaggtaaaaaatcaagtgagggtggaagctccggtaaagggt<br>ggtggtttgggaaaaagcggagaatccggtaaagggtgatcaagt<br>ataggggtattggtggaacccggcataaatcaaggcgataat<br>aatgatacttcaaatcaaacgctgatagtagaagaaaattcgactg<br>tgaataatgcagatcagactaagagtgtagaaaaaacaaagggt<br>gatcaaaaatctgaacaaaataattcaaatgaagtaaatactactg<br>aagggtgtcagaacaaatccggaatgactaataatacaagtgaat<br>atactataataaacctactgtgaataaagaaaaatgaagatgaaaa<br>aactaatacaacacaggataacaaagatcggatagcacttcaaa<br>tgatttgagtaaaaacaaaagatacacagacaaatacaaatgatag<br>tgatataaagtcggctactgcttcaaatgccgaatccaaaaggaa<br>gaagaacataaaaatacagaactacgaatagttctgagaagaa<br>taaaaatactgaacagcagaacaaatctgaagagaatatagtag<br>agagtgatcaggataaggccaatagaataaaatcctcttgactac<br>cgttaaaaatgataatgtgttacagtacacgaatgaagatggaga<br>aattagaacaattatttggcacaagggtcatcaatggacctactatg<br>ggagagggtggagattttcacaagaaataacaaatgcgggttca<br>gatacatatagattataaaacttcgttactacagggtgatggatgg<br>tatgatgtaaataagacaagagccggtggcaatgtagacgtatgat<br>aaaaatcttgccttggggcagtttcatccaatatgtacattggtggtt<br>gaacaaaatattgaatatattgacagatatatagtaaagaatggag<br>ataataatctgtgcaaataaggaagatctcagatctgaaatattctttaa<br>gagtcagcaaaaatagtgctatattcgaactcttaagggttatttggga<br>aataatgagaagggtattttatacagacctttgatggacctgtttataa<br>acggatatactccaagacagggtggaggtaacaaatatagaaga<br>gatgatttaacacctgatagcagagggtgattttctatgatgtgttaa<br>aggagagaagctgacagaccgtacttatggtggaagttatgaagc<br>tttaggtgataaattaaaagaagtactgggagaagccggtttgatag<br>gtatatctcataaggcttttaaaaacagtaatcatatagtaactctatg<br>gggtgctgaatatgatttgaatgaaagcttatagcggatatatattagt<br>gattcagatgatcaaaatgaaacagaagatgaaagagatctggg<br>catgaaacgctatgaggtaggaatgtaggaggagtggcaaaact<br>gtccacaaatcaaacgaataagtcgtcagggttcagctataggatat<br>cttcatatactatacttaggaaccacacagtgggaaaattattttaaataa | SEQ ID NO: 2 |
| SEQ ID NO:<br>14 | atgggcatcgaagggtggaaattcgagggtacaactggaagtac<br>tataaagaggataaggcgctgaaagcgtggcagcaactgaaca<br>acaaatgggtacttcttaacgaagacggcagcctgaagaccggttg<br>gcacctggatgcgaccaacaactggtatttctggacagcagcaa                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | SEQ ID NO: 3 |

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |              |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
|                  | <p> a a a c a g c a a c g a g g g c a t g c t g c t g a g c g g t t g c a a t g g a t c g<br/> a t g g t a c t g c t a c t a c t t c g a a a a c a t c g a c a a g a g c a c c c t g g g<br/> c c g t a t g t a c g t t a a c a c c g t g g t a a c g g c t a t c g t a c g a c g c g a<br/> c c g g t c g t t g g a c c a a c g a g a a a g g c g t t g a a t a c t a t g a g g c g g<br/> g c a a g g g t a t c a t t a c c g t g a a c a c c c t g c a g g g c a g c c t g a a c a<br/> c c g g t a a g a a a g c a g c g g t g g c g g t a g c a g c g g c a a g g g c g<br/> g t g g c c t g g g c a a a a g c g g t g a a a g c g g c a a g g g t g g c a g c a g<br/> c a t c g g t g g c a t t g g t g g c a a a a c c g g c a t c a a c c a g g g t g a c a<br/> a c a a c g a t a c c a g c a a c c a a a g c c t g a t t g t g a g g a a a c a g c<br/> a c c g t g a a c a a c g c g g a t c a a a c c a a a a g c g t t g a g a a g a c c a<br/> a a g g t g a c c a g a a g a g c g a a c a a a a c a c a g c a a c g a a g t g a<br/> a c a c c a c c g a g g g t g g c c a g a a c a a g a g c g g c a t g a c c a a c a<br/> a c a c c a g c g a g t a c a c c t a t a c a a g c g a c c g t t a c a a a g a a<br/> a a c g a g g a t g a a a a a c c a a c a c c a c c a g g a c a c a a a g t g<br/> a g c g a t a g c a c c a g c a a c g a c c t g a g c a a g a c c a a a g a c a c c<br/> c a a c c a a c a c c a a c g a c a g c g a t a t c a a a a g c g c g a c c g c g<br/> a g c a a c g c g g a g a t t c a g a a g g a a g a g g a g c a c a a a a c a c c<br/> g a a a c c a c a a c a g c a g c g a g a a g a c a a a a a c a c c g a a c a<br/> g c a a a c a a g a g c g a a g a g a a c a t c g t t g a a a g c g a c c a a g a t<br/> a a g g c g a a c c g t a t t a a a g c a g c c t g a c c a c c g t g a a a a c g<br/> a t a a c g t g g t c a g t a c c a a c g a a g a c g g c g a g a t c c g t a c c<br/> a t c a t t t g g g c g a a g g g t a t a a c g g c c c g a c c a t g g g c g a g g g t<br/> g g c g a t t t c c a a a g a a a t c a c c a a c g c g g g c a g c g a c a c c t a<br/> c a t t g a t t a t a a g a c c a g c t t a c c a c c g g c g a c g g t t g g t a t g a t<br/> t a a c a a a a c c c g t g c g g t g g c a a c g t g g a c g t t g a t a a g a a c c t<br/> g t g c t t c g g t g c t g t g a g c a g c a a c a t g t g c a c t g g t g g t t g a a c<br/> a g a a c a t c g a g t a c a t t g a t c g t t a t a t c g t t a a a a c g g c g a c a t c<br/> a t t c g t g c g a a c c g t a a g a t t a g c g a c c t g a a g t a c a g c t t c a a a a<br/> g c c a g c a a a a c a g c g c g a t c t c g a g c t g t t a a a g t g c t g t t c g g<br/> c a a c a a c g a a a a g g g t t t t a c a c c g a c c t g c t g a t g g a c c t g t t c<br/> a t c a a c g g t t a t a c c c c g a a a a c c g g t g g c g g t a c c a a c a t t g a g<br/> c g t g a c g a t c t g a c c c c g g a t a g c c g t g g c g g t t c t t t a c a c g t t t<br/> t c a a g g g c g a g a a a c t g a c c g a t c g t a c c t a c g g c g g t a g c t a t g<br/> a a g c g t g g g c g a a a g c t g a a a g a a g t g c t g g g c g a g g c g g g<br/> c c t g a t c g g t a t t a g c c a a a g g c g t t a a a a t a g c a a c c a c a t c<br/> g t t a c c c t g t g g g t g c g g a g t a c g a t c t g a a c g g c a a a c t g a t c g<br/> c g g t g a t a t t a g c g a c a g c g a c g a t c a g a a c g a g a c c g a a g a c<br/> g a g c g t g a t c t g g g t a t g a a g c g t t a c g a a g t g c g t a a c g t t g g c g<br/> g t g t g g c g a a a c t g a c c a c c a a a c c a a c a a g a g c a g c g<br/> g c a g c g c g a t c g g t a c c t g c a c a t t c t g a t c t g g g t a c c a c c c a g<br/> t g g g a a a c t a t t c a a g c t c g a g c a c c a c c a c c a c c a c c a c t g a </p> |              |
| SEQ ID NO:<br>15 | <p> a t g c a c c a c c a c c a c c a c c a c c a c c a c c a c c a c c g g c a t c g a a g<br/> g t t g g a a a t t c a g g g t a a c a a c t g g a a g t a c t a t a a a g a g g a t a<br/> a g g c g t g a a a g c g t g g c a g c a a c t g a a c a c a a a t g t a c t t c t<br/> t t a a c g a a g a c g g c a g c c t g a a g a c c g g t t g g a c c t g g a t g c g<br/> a c c a a c a a c t g g t a t t t c t g g a c a g c a g c a a a a c a g c a a c g a<br/> g g g c a t g c t g t g a g c g g t t g g c a a t g g a t c g a t g g t a c t g t a c t<br/> a c t t c g a a a c a t c g a a a g a g c a c c c t g g g c c g t a t g t a c g t t a<br/> a c a c c g t g g t a a c g g c t a t c g t a t c g a c g c g a c c g g t c g t t g g a c<br/> c a a c g a g a a a g g c g t t g a a t a c t a t g a g g c g g g c a a g g g t a t c a<br/> t t a c c g t g a a c a c c c t g a g g g c a g c c t g a a c a c c g g t a a g a a a<br/> a g c a g c g g t g g c g g t a g c a g c g g c a a g g g c g g t g g c c t g g g c a<br/> a a a g c g g t g a a a g c g g c a a g g g t g g c a g c a g c a t c g g t g g c a t t<br/> g g t g g c a a a a c c g g c a t c a a c c a g g g t g a c a a c a c g a t a c c a<br/> g c a a c c a a a g c c t g a t t g t g a g g a a a c a g c a c c g t g a a c a a c </p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | SEQ ID NO: 4 |

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |              |
|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
|                  | <p>gcggatcaaaccaaaagcgttgagaagaccaaaggtgaccaga<br/> agagcgaaacaaaacaacagcaacgaagtgaacaccaccgag<br/> ggtggccagaacaagagcggcatgaccaacaacaccagcgag<br/> tacacctataacaagccgaccgttaacaaagaaaacgaggatga<br/> aaaaaccaacaccaccaggacaacaaagttagcgatagcac<br/> cagcaacgacctgagcaagaccaaagacacccaaaccaacac<br/> caacgacgagcgtatcaaaagcgcgaccgagcaacgcgga<br/> gattcagaaggaagaggagcaaaaacaccgaaaccaccaa<br/> cagcagcgagaagaacaaaaacaccgaacagcaaaacaaga<br/> gcgaagagaacatcgttgaaagcgaccaagataaggcgaaccg<br/> tattaaaagcagcctgaccaccgtgaaaaacgataacgtggtca<br/> gtaccaacgaagacggcgagatccgtaccatcattgggcga<br/> agggtattaacggcccgaccatggcgagggtggcgattccaca<br/> aagaaatcaccaacgcgggcagcgacacctacattgattataag<br/> accagcttaccaccggcgacggttggtatgatgtaacaaaaccc<br/> gtgcgggtggcaacgtggacgttgataagaacctgtgctcggtgct<br/> gtgagcagcaacatgctgcactggtggttgaaacagaacatcgagt<br/> acattgatcgttatatcgtaaaaacggcgacatcattcgtgcgaac<br/> cgtaagattagcgacctgaagtacagcttcaaaagccagcaaaa<br/> cagcgcgatcttcgagctgttaagtgctgttcggcaacaacgaa<br/> aagggttttacaccgacctgctgatggacctgtcatcaacggtat<br/> accccgaaaacccggtggcggtaccaacattgagcgtgacgatctg<br/> accccgatagccgtggcggttctttacgacgtttcaaggcgga<br/> gaaactgaccgatcgtacctacggcggtagctatgaagcgctggg<br/> cgacaagctgaaagaagtgcgtggcgaggcgggcctgatcggtat<br/> ttagccacaaggcggttaaaaatagcaaccacatcgttaccctgtg<br/> gggtgcggagtacgatctgaacggcaaacgatcgcggtgtatatt<br/> agcgacagcgacgatcagaacgagaccgaagacgagcgtgat<br/> ctgggtatgaagcgttacgaagtgcgtaacgttgcggtgtggcga<br/> aactgagcaccaaccaaccaacaagagcagcggcgagcgga<br/> tcggttacctgcacattctgtatctgggtaccaccagtgggaaaact<br/> attcaagctcgagtga</p> |              |
| SEQ ID NO:<br>16 | <p>atgcagaacaagagcggcatgaccaacaacaccagcgagtac<br/> acctataacaagccgaccgttaacaaagaaaacgaggatgaaa<br/> aaaccaacaccaccaggacaacaaagttagcgatagcacca<br/> gcaacgacctgagcaagaccaaagacacccaaaccaacacca<br/> acgacagcgatataaaaagcgcgaccgagcaacgcgggaga<br/> ttcagaaggaagaggagcaaaaaacaccgaaaccaccaaca<br/> gcagcgagaagaacaaaaacaccgaacagcaaaacaagagc<br/> gaagagaacatcgttgaaagcgaccaagataaggcgaaccgtat<br/> taaaagcagcctgaccaccgtgaaaaacgataacgtggttcagta<br/> caccaacgaagacggcgagatccgtaccatcattgggcgaagg<br/> gtattaacggcccgaccatggcgagggtggcgatttcacaaag<br/> aatcaccaacgcggcgagcgacacctacattgattataagacca<br/> gctttaccaccggcgacggttggtatgatgttaacaaaacccgtgc<br/> gggtggcaacgtggacgttgataagaacctgtgctcggtgctgtga<br/> gcagcaacatgctgcactggtggttgaaacagaacatcgagtacat<br/> tgatcgttatatcgtaaaaacggcgacatcattcgtgcgaaccgta<br/> agattagcgacctgaagtacagcttcaaaagccagcaaaacagc<br/> gcgatcttcgagctgttaagtgctgttcggcaacaacgaaaagg<br/> gttttacaccgacctgctgatggacctgttcatcaacggttatcccc<br/> gaaaaccggtggcggtaccaacattgagcgtgacgatctgacccc<br/> ggatagccgtggcggttctttacgacgtttcaaggcgagaaact<br/> gaccgatcgtacctacggcggtagctatgaagcgctggcgaca<br/> agctgaaagaagtgcggcgaggcggcgctgatcggtattagcc</p>                                                                                                                                                                                                                                                                                                                                                                                                                                 | SEQ ID NO: 5 |

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |              |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
|                  | acaaggcggttaaaaatagcaaccacatcggtaccctgtggggtgc<br>ggagtacgatctgaacggcaaaactgatcgcggtgtatattagcgac<br>agcgacgatcagaacgagaccgaagacgagcgtgatctgggtat<br>gaagcggttacgaagtgcgtaacgttggcggtgtggcgaaactgag<br>caccaaccaaccaacaagagcagcggcagcgcgatcggttac<br>ctgcacattctgtatctgggtaccaccagtgggaaaactatttcaa<br>gctcgagcaccaccaccaccactga                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |              |
| SEQ ID NO:<br>17 | atgagcgatagcaccagcaacgacctgagcaagaccaaagac<br>acccaaaccaacaccaacgacagcgatataaaagcgcgacc<br>gcgagcaacgaggagattcagaaggaagaggagcacaaaaac<br>accgaaaccaccaacgacgagcagaagaacaaaaacaccga<br>acagcaaaaacaagagcgaagagaacatcgttgaagcgacca<br>agataaggcgaaccgtattaaaagcagcctgaccaccgtgaaaa<br>acgataacgtggtcagtagaccaacgaagacggcgagatccgt<br>accatcatttggcggaagggtattaacggcccaccatggcgag<br>ggtggcgatttcacaaaagaaatcaccaacgcgggcagcgacac<br>ctacattgattataagaccagctttaccaccggcgacggttggtatg<br>atgttaacaaaaccgtgcgggtggcaacgtggacgttgataaga<br>acctgtgcttcggtgctgtgagcagcaacatgctgcactggtggttg<br>aacagaacatcgagtacattgatcgttatatcgttaaaaacggcga<br>catcattcgtgcgaaccgtaagattagcgacctgaagtacagcttc<br>aaaagccagcaaaacagcgcatcttcgagctgtttaaagtgtgt<br>tcggcaacaacgaaaagggttttacaccgacctgctgatggacct<br>gttcatcaacggttatacccgaaaaccggtggcggtaccaacatt<br>gagcgtgacgatctgaccccgatagccgtggcggttctttacga<br>cgtttcaaggcgagaaaactgaccgatcgtacctacggcggttag<br>ctatgaagcgtggcgacaagctgaaagaagtgtggcgagg<br>cgggcctgatcgggtattagccacaaggcggttaaaaatagcaacc<br>acatcgttaccctgtgggtgcggagtacgatctgaacggcaaaact<br>gatcgcggtgtatattagcgacagcgacgatcagaacgagaccg<br>aagacgagcgtgatctgggtatgaagcgttacgaagtgcgtaacg<br>ttggcggtgtggcgaaactgagcaccaaccaaccaacaagag<br>cagcggcagcgcatcggttacctgcacattctgtatctgggtacca<br>cccagtgggaaaactatttcaagctcgagcaccaccaccacc<br>actga | SEQ ID NO: 6 |
| SEQ ID NO:<br>18 | atgaacatcgttgaaagcgaccaagataaggcgaaaccgtattaaa<br>agcagcctgaccaccgtgaaaaacgataacgtggtcagtagcacc<br>aacgaagacggcgagatccgtaccatcatttggcggaagggtatt<br>aacggcccagaccatgggcgaggttgcggtttccacaaagaaat<br>caccaacgcgggcagcgacacctacattgattataagaccagctt<br>accaccggcgacggttggtatgatgttaacaaaaccgtgcgggt<br>ggcaacgtggacgttgataagaacctgtgcttcggtgctgtgagca<br>gcaacatgctgcactggtggttgaacagaacatcgagtacattgat<br>cgttatatcgttaaaaacggcgacatcattcgtgcgaaccgtaagat<br>tagcgacctgaagtacagcttcaaaagccagcaaaacagcgcg<br>atcttcgagctgtttaaagtgtgttcggcaacaacgaaaagggtttt<br>acaccgacctgctgatggacctgtcatcaacggttatacccgaa<br>aaccggtggcggtaccaacattgagcgtgacgatctgacccgga<br>tagccgtggcggttctttacgacgtttcaaggcgagaaaactgac<br>cgatcgtacctacggcggtagctatgaagcgtggcgacaagct<br>gaaagaagtgtggcgagggcggtcgtatcggtattagccaca<br>aggcggttaaaaatagcaaccacatcgttacctgtgggtgcgga<br>gtacgatcgaacggcaaaactgatcgggtgtatattagcgacagc<br>gacgatcagaacgagaccgaagacgagcgtgatctgggtatgaa<br>gcgttacgaagtgcgtaacgttggcggtgtggcgaaactgagcac                                                                                                                                                                                                                                                                                                                                          | SEQ ID NO: 7 |

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |              |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
|                  | caaccaaaccaacaagagcagcggcagcgcgatcggttacctg<br>cacattctgtatctgggtaccaccagtgaggaaaactatttcaagctc<br>gagcaccaccaccaccactga                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |              |
| SEQ ID NO:<br>19 | gcagaaagaaaagagagtacacaaaacagggctgagagattta<br>aagaatcattaattacttcagagaatgacaatgttggtcagtatacaa<br>atgaagacggtgaaatacgaacctcatatgggtaaagggaatta<br>acggacctgttatgggagaagggtggagactttcataaagagattac<br>acaaggcggtaagaatacttatgtaatgtatactgtaccatattcaa<br>gcgagacggatggtatgatgtaataaaacaagaagcgggtgga<br>aatatcgatatagataaaaatcttgccttgacgaggatcatcaaat<br>atgcttcattggtggttgatcaaaatattgaatatgtgacagatatat<br>agaaaaaacggagatataattcgtgccaacagaaaactttcag<br>atttaaaaacttcttgaaaaatccggaaaacagtaaaatatttgag<br>ctgtataaagtattgtacggatataatgagagagggttttattcggatc<br>tttgatggatttttatttaacggatatactccaacccaaagggtgg<br>aacaatatagaaagtgtatcttatcccaataataacggcgg<br>attcttctatgatgtttcaaggggaaaaaacttacagaccgtacttat<br>ggtggaaactatgaatcactaagtaattctttaaagaaatctgggt<br>aatggcggcatagtaggactttctataaagtgtttcaagagtaaat<br>catatagtactttgtgggtgccgaatatgactgaacgggaagtgtg<br>aaggctgtatatataagcgattctgacgatcaggatgaaagtaagt<br>aggcatgaaacgtttgaaatccgaaatgtaggaggaaaagcaa<br>aaatatctaccaacgaaacagataagtctcgggtgctgaattagg<br>atatctcatatcctatatcagggaacccggacaatgggaaaattttt<br>aaataa                                                                                                                                                                                                                                                                                                                                                                                                                             | SEQ ID NO: 8 |
| SEQ ID NO:<br>20 | atgaaaaatataaagtcggctctgtattgtgggactactatgttga<br>tgtccttaatgcctatgcagggtgtgatggatggaacaggaaaat<br>aacagttggaatattataaagaagataaggctctaaggcttgcc<br>agcagataaataataattggtatttcttaatgaagacggaagttaa<br>aaaccggttggtatatggacgccgcaaataattggtatttttagaca<br>gctctaaaactgccaatgagggagtgttttatcaggatggcagtgg<br>atagacggatattgctattttgagggaatagataataatacattcg<br>gaaagatgtattcaatactctgtcagcggatataaagtgtatgctt<br>aggaagatggattaatgaaagcggagcagaatattatgaagccg<br>gtaaaggataagtagcaggtaatgcttcacaaagttcatcgaatac<br>agcaagaaaaacaagtgtgcagggaagttcggaaaaagcgggt<br>aaaagcggaggctcaagcaaggcgggaagctccgaaaaagc<br>ggaggctcaagcaaggcgggaagtccggtaaaagcggaggct<br>caagcttttaaatcaaggagaaaaagagaagaaaaacagaata<br>ttaggagaatcaaaaggaaaggctgaaaactcaagttccaaaag<br>aagtgaagacttggtagggtcaagacaagacttaagctcggataa<br>aaataacagcttgattgcagataataaaacagggaatagtggaa<br>agccggaagcatagaaaaaaatacaataaagagtataaccaca<br>aagtcaaatgagagccctgtaaaaccggccacttctcaaatgtcg<br>agataccgaaagaagaggaaaaaaaggatatagagtcaaagg<br>aaagtaaaaagcctgaggcagaaagaaaagagagtacacaaa<br>acagggtgagagatttaagaatcattaattacttcagagaatga<br>caatgtgttcagtatacaaatgaagacggtgaaatacgaacctatc<br>atatgggttaaagggaattaacggacctgttatgggagaagggtgga<br>gactttcataaagagattacacaaggcggtaagaatacttatgtaat<br>gtatactgtaccatattcaagcggagacggatggtatgatgtaata<br>aaacaagaagcgggtgaaatatcgatatagataaaaatcttgcctt<br>gcagcggtatcatcaaatatgcttcattggtggttgatcaaaatattg<br>aatatgttgacagatatatagaaaaaacggagatataattcgtgc<br>caacagaaaactttcagatttaaaaacttcttgaaaatccggaaa | SEQ ID NO: 9 |

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |               |
|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
|                  | acagtaaaatatttgagctgtataaagtattgtacggatataatgaga<br>gaggggtttattcggatctttgatggattattattaacggatatactcc<br>aaacccaaagggtggaacaaatagaaagtgatgatcttatccc<br>aaataataacggcggattcttctatgatgtttcaaggggaaaaaac<br>ttacagaccgtacttatggtggaactatgaatcactaagtaatcttct<br>taaagaaatcttgggtaatggcggcatagtaggactttctcataaag<br>tgtttcaagagtaatcatatagttactttgtgggtgccgaatatga<br>ctgaacgggaagttgaaggctgtatatataagcgattctgacgatc<br>aggatgaaagtaatgtaggcatgaaacgtttgaaatccgaaatgt<br>aggaggaaaagcaaaaatatctaccaacgaaacagataaagtct<br>cggggtgctgaattaggatattctcatatcctatatcagggaaacgg<br>acaatgggaaaattatttaataa                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |               |
| SEQ ID NO:<br>21 | atgggtgtggacggctggaacaggagaacaacagctggaagt<br>actataaagaagataaggcgtgaaagcgtggcagcaaatcaac<br>aacaactggtatttcttaacgaggacggtagcctgaaaacggct<br>ggtacatggacgcgcggaacaactggtatttctggatagcagca<br>agaccggaacgaggggttctgctgagcggctggcagtggaatcg<br>acggttactgtactatttgaaggcattgataacaacacctttgta<br>aaatgtacagcaacaccctggttagcggctataaggtgacgcgct<br>gggtcgttgatcaacgagagcggcggaatactatgaggcgg<br>gtaaaggcattagcaccggtaacgcgagcaaagcagcagcaa<br>caccgcgaagaaaaccagcgtggcgggttagcagcgaagag<br>cggcaaaagcgggtggcagcagcaagggtggcagcagcggtaa<br>gagcgggtggcagcagcaagggtggcagcagcgggaaaagcgg<br>tggcagcagcttctgaaccagggcgagaagcgtgaggaaaac<br>cgtatcctgggtgaaagcaagggcaaagcggagaacagcagca<br>gcaaacgtagcgaagacctggttggtagccgtcaagacctgagc<br>agcgataagaacaacagcctgatcgcggaatacaagaccggt<br>acagcgaaaaagcgggcagcattgagaagaacaccaacaag<br>aatataccaccaagagcaacgagagcccgtgaaaccggcgac<br>cagcagcaacgttgaatcccgaaggaagaggagaagaagga<br>cattgaaagcaagagagcaagaaaccggaagcggagcgtaa<br>ggagagcaccagaaccgtgcggaacgttcaaagagagcctg<br>atcaccagcgaaaacgacaacgtggttaatataccaacgaaga<br>tggcgagatccgtaccatcatttgggtgaagggtattaacggcccg<br>gttatggcgagggtggcgatttcaacaaggaaattaccaggggtg<br>gcaaaaacacctacgtgatgtataccgttccgtacagcagcgggtg<br>acggctggtatgatgtgaacaaaaccgtagcgggtggcaacatcg<br>acattgataagaacctgtgcttcgcggtgtgagcagcaacatgct<br>gcactggtggttgacaaaacatcgagtacgttgatcgttatattga<br>gaagaacggtgacatcattcgtgcgaaccgtaagctgagcgtatc<br>gaaaaccagcttcgaaaaccgggagaacagcaagatctttgagc<br>tgtacaaagtgtgtacgggtataacgaacgtggcttctatagcga<br>ctgctgatggacctgttcaaacggttacaccccgaacccaagg<br>gtggcaccaacatcgagagcgacgatctgattccgaacaacaac<br>gggtggcttctttacgacgttttcaaaggcaagaaactgaccgatcgt<br>acctcgggtggcaactatgagagcctgagcaacctgtgaaggga<br>aatcctgggtaacggtggcattgtggcctgagccacaaggtttta<br>gcaaaagcaaccacatcgtgacctgtgggtgaggagtagcagc<br>ctgaacggcaagctgaaagcgggttatattagcgacagcgacgat<br>caagatgagagcaacgtgggtatgaaacgtttcgaatccgtaac<br>gttgggtggcaaggcgaatagcaccacgagaccgataagag<br>cgcggtgcggaactgggtacctgcacatcctgtatcagggtacc<br>ggccaatgggaaaactactttaagctcagcaccaccaccacca<br>ccactga | SEQ ID NO: 10 |

|               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |               |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| SEQ ID NO: 22 | atgcaccaccaccaccaccaccaccaccacggtgtgagcg<br>gctggaacacaggagaacaacacgctggaagtactataaagaaga<br>taagcgctgtaaagcgtggcgagcaaatcaacaacaactggtattt<br>ctttaacgaggacggtagcctgaaaaccggtggtacatggacgc<br>ggcgaacaactggtatttctgtagcagcaagaccgcgaacg<br>aggggtgttctgctgagcggctggcagtgatcgacggttactgtac<br>tatttgaaggcattgataacaacacctttggtaaaatgtacagcaa<br>caccctggttagcggctataagggtgacgcgctgggtcgttgatca<br>acgagagcggcggaatactatgaggcggttaaaggcattagc<br>accggtaacgcgagccaaagcagcagcaacaccgcgaagaaa<br>accagcgtggcggttagcagcgaagagcggcaaaagcgggt<br>ggcagcagcaaggggtggcagcagcggtaagagcgggtggcagc<br>agcaaggggtggcagcagcgggaaaagcgggtggcagcagcttc<br>tgaaccagggcgagaagcgtgaggaaaaccgtatcctgggtgaa<br>agcaagggcgaagcggagaacagcagcagcaaacgtagcga<br>agacctggttgtagcgtcaagacctgagcagcgataagaaca<br>acagcctgatcgcgataacaagaccggtaacagcgaaaaagc<br>gggcagcattgagaagaacaccaacaagaatataccaccaag<br>agcaacgagagcccgtgaaaccggcgaccagcagcaacgttg<br>aatcccgaaggagaggagaagaaggacattgaaagcaaag<br>agagcaagaacccggaagcggagcgtaaggagagcaccag<br>aaccgtgcggaacgtttcaaagagagcctgatcaccagcgaaaa<br>cgacaacgtggttaataataccaacgaagatggcgagatccgtac<br>catcattgggtgaagggatataacggcccgggtatggcgaggggt<br>ggcgattttacaaggaaattaccaggggtggcaaaaacacctac<br>gtgatgtataccgttccgtacagcagcgggtgacggctggtatgatgt<br>gaacaaaaccgtagcgggtggcaacatcgacattgataagaacc<br>tgtgtctgcggctgtgagcagcaacatgctgactggtggttgac<br>caaaacatcgagtacgttgatcgttatattgagaagaacgggtgaca<br>tcattcgtgcgaaccgtaagctgagcgtatcgaaccagcttcga<br>aaaccgggagaacagcaagatcttgagctgtacaagtgtgtga<br>cgggtataacgaacgtggcttctatagcgacctgctgatggacctgtt<br>cattaacgggttacaccccgaaccgaagggtggcaccacatcg<br>agagcgcagatctgattccgaacaacaacgggtggcttctttacga<br>cgttttcaaagggaagaaactgaccgatcgtacctacgggtggcaa<br>ctatgagagcctgagcaacctgctgaaggaaatcctgggtaacgg<br>tggcattgtggcctgagccacaaggttttagcaaaagcaaccac<br>atcgtgacctgtgggtgctggagtagcagctgaacggcaagctg<br>aaagcgggttatattagcgacagcgacgatcaagatgagagcaac<br>gtgggtatgaaacgtttcgaatccgtaacgttggtggcaaggcga<br>aaattagcaccaacgagaccgataagagcgcgggtgcggaact<br>gggtacctgcacatcctgtatcagggtaccggccaatgggaaaa<br>ctactttaagctcgaqtga | SEQ ID NO: 11 |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|

It will be understood that any number of polynucleotide sequences may be utilized to encode a given polypeptide according to the invention due to the degeneracy of the genetic code. As such, the polynucleotide sequences described herein are illustrative only.

## Polypeptides

The present invention relates to polypeptides having IgM specific endoprotease activity. In preferred embodiments, the IgM specific endoprotease activity is human IgM specific endoprotease activity.

5

In one aspect, there is provided a polypeptide having endoprotease activity comprising:

- d) the amino acid sequence of SEQ ID NO: 1;
- e) a variant thereof having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 1; or
- 10 f) a fragment of either a) or b),

preferably wherein said endoprotease activity is human IgM specific endoprotease activity.

In another aspect, there is provided a polypeptide having endoprotease activity comprising:

- d) the amino acid sequence of SEQ ID NO: 8;
- 15 e) a variant thereof having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 8; or
- f) a fragment of either a) or b),

preferably wherein said endoprotease activity is human IgM specific endoprotease activity.

- 20 In one embodiment, the polypeptide of the invention has IgM specific and IgG specific protease activity and said polypeptide comprises or consists of a sequence of any one SEQ ID NOs: 8 to 11, or a variant or fragment thereof, as defined herein.

In some embodiments, the IgG specific activity is lower than the IgM specific activity.

25

In an embodiment of the invention, there is provided a polypeptide having IgM specific endoprotease activity, wherein the polypeptide comprises or consists of a polypeptide selected from the group consisting of SEQ ID NOs: 1 to 7.

- 30 In an embodiment of the invention, there is provided a polypeptide having IgM specific endoprotease activity, wherein the polypeptide comprises or consists of a polypeptide having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1 to 7.

In an embodiment of the invention, there is provided a polypeptide having IgM specific endoprotease activity, wherein the polypeptide comprises or consists of a polypeptide having at least 80% sequence identity, such as at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least  
5 97% sequence identity, at least 98% sequence identity, at least 99% sequence identity, to a sequence selected from the group consisting of SEQ ID NOs: 1 to 7.

In an embodiment of the invention, there is provided a polypeptide having IgM specific endoprotease activity, wherein the polypeptide comprises or consists of a polypeptide having  
10 at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 8 to 11.

In an embodiment of the invention, there is provided a polypeptide having IgM specific endoprotease activity, wherein the polypeptide comprises or consists of a polypeptide having  
15 at least 80% sequence identity, such as at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least 97% sequence identity, at least 98% sequence identity, at least 99% sequence identity, to a sequence selected from the group consisting of SEQ ID NOs: 8 to 11.

20 The present invention also encompasses fragments of other polypeptides of the invention. Said fragments are to be considered polypeptides in their own right.

In other words, polypeptides of the invention may comprise or consist of fragments or shorter sections of longer polypeptides according to the invention, as defined herein.

25 A fragment will be understood to be any continuous polypeptide sequence that is shorter than a defined longer sequence with which it is identical (across the amino acids that they have in common). For example, a polypeptide that is 200 amino acids in length may be considered a fragment of a longer polypeptide sequence of 250 amino acids in length with  
30 which it shares 100% sequence identity.

In one embodiment, the polypeptide comprises or consists of a fragment of a polypeptide sequence selected from the group consisting of SEQ ID NOs: 1 to 7.

35 In one embodiment, the polypeptide comprises or consists of a fragment of a polypeptide sequence selected from the group consisting of SEQ ID NOs: 8 to 11.

In one embodiment, the polypeptide comprises or consists of a fragment of a polypeptide sequence selected from the group consisting of SEQ ID NOs: 1 to 11, wherein the fragment is at least 100 amino acids in length, such as at least 150 amino acids in length, at least 200 amino acids in length, at least 250 amino acids in length, at least 300 amino acids in length, at least 350 amino acids in length, at least 400 amino acids in length, at least 450 amino acids in length, at least 500 amino acids in length, at least 550 amino acids in length, at least 600 amino acids in length, or at least 650 amino acids in length.

In one embodiment, the polypeptide comprises or consists of a fragment of a polypeptide sequence selected from the group consisting of SEQ ID NOs: 1 to 11, wherein the fragment is at least 300 amino acids in length, such as at least 310 amino acids in length, at least 320 amino acids in length, at least 330 amino acids in length, at least 340 amino acids in length, at least 350 amino acids in length, at least 360 amino acids in length, at least 370 amino acids in length, at least 380 amino acids in length, at least 390 amino acids in length, or at least 400 amino acids in length.

The polypeptide is typically at least 75, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 500, at least 550, at least 600, or at least 650 amino acids in length.

In preferred embodiments, the polypeptide is at least 300 amino acids in length, such as at least 310 amino acids in length, at least 320 amino acids in length, at least 330 amino acids in length, at least 340 amino acids in length, at least 350 amino acids in length, at least 360 amino acids in length, at least 370 amino acids in length, at least 380 amino acids in length, at least 390 amino acids in length, or at least 400 amino acids in length.

The polypeptide may be no longer than 700 amino acids in length, such as no longer than 650, no longer than 600, no longer than 550, no longer than 500, no longer than 450, no longer than 400, no longer than 350 no longer than 300, no longer than 250, no longer than 200, no longer than 150, or no longer than 100 amino acids in length.

It will be appreciated that any of the above listed lower limits may be combined with any of the above listed upper limits to provide a range for the length the polypeptide. For example, the polypeptide may be 75 to 700 amino acids in length, or 450 to 550 amino acids in length.

In one embodiment the polypeptide is between 300 and 700 amino acids in length, such as between 310 and 690 amino acids in length, 320 and 690 amino acids in length, 330 and

680 amino acids in length, 330 and 670 amino acids in length, 320 and 640 amino acids in length, and 320 and 630 amino acids in length.

In a preferred embodiment the polypeptide is between 300 and 700 amino acids in length,  
5 more preferably between 320 and 690 amino acids in length.

In one embodiment, there is provided a polypeptide having endoprotease activity comprising or consisting of the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

10 In one embodiment, there is provided a polypeptide having endoprotease activity comprising or consisting of a polypeptide having at least 80% sequence identity, such as 85% sequence identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11..

15

In one embodiment, there is provided a fragment of polypeptide, the fragment having endoprotease activity, wherein the fragment comprises or consists of a fragment of the amino acid sequence of any one of SEQ ID NOs: 1 to 11, or a fragment of an amino acid sequence having at least 80% sequence identity, such as 85% sequence identity, 90%  
20 sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

In one embodiment, there is provided a polypeptide having endoprotease activity, the  
25 polypeptide comprising:

- a) the amino acid sequence of any one of SEQ ID NOs: 1 to 11;
- b) a variant thereof having at least 80% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11; or
- c) a fragment of either a) or b).

30

In one embodiment, there is provided a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid sequence having at least 80% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

35

In one embodiment, there is provided a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid

sequence having at least 85% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

5 In one embodiment, there is provided a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid sequence having at least 90% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

10 In one embodiment, there is provided a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid sequence having at least 95% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

15 In one embodiment, there is provided a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid sequence having at least 98% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

20 In one embodiment, there is provided a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid sequence having at least 99% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

25 In one embodiment, there is provided a polypeptide having endoprotease activity comprising or consisting of the amino acid sequence of SEQ ID NO: 1.

30 In one embodiment, there is provided a polypeptide having endoprotease activity comprising or consisting of a polypeptide having at least 80% sequence identity, such as 85% sequence identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of SEQ ID NO: 1.

35 In one embodiment, there is provided a fragment of polypeptide, the fragment having endoprotease activity, wherein the fragment comprises or consists of a fragment of the amino acid sequence of SEQ ID NO: 1, or a fragment of an amino acid sequence having at least 80% sequence identity, such as 85% sequence identity, 90% sequence identity, 95%

sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of SEQ ID NO: 1.

In one embodiment, there is provided a polypeptide having endoprotease activity, the polypeptide comprising:

- a) the amino acid sequence of SEQ ID NO: 1;
- b) a variant thereof having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 1; or
- c) a fragment of either a) or b).

In one embodiment, there is provided a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 1.

In one embodiment, there is provided a polypeptide having endoprotease activity comprising the amino acid sequence of SEQ ID NO: 7.

In one embodiment, there is provided a polypeptide having endoprotease activity consisting of the amino acid sequence of SEQ ID NO: 7.

In one embodiment, there is provided a polypeptide having endoprotease activity comprising or consisting of a polypeptide having at least 80% sequence identity, such as 85% sequence identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of SEQ ID NO: 7.

In one embodiment, there is provided a fragment of polypeptide, the fragment having endoprotease activity, wherein the fragment comprises or consists of a fragment of the amino acid sequence of SEQ ID NO: 1, or a fragment of an amino acid sequence having at least 80% sequence identity, such as 85% sequence identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of SEQ ID NO: 7.

In one embodiment, there is provided a polypeptide having endoprotease activity, the polypeptide comprising:

- d) the amino acid sequence of SEQ ID NO: 7;

- e) a variant thereof having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 7; or
- f) a fragment of either a) or b).

5 In one embodiment, there is provided a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 7.

10 In one embodiment, there is provided a polypeptide having endoprotease activity comprising or consisting of the amino acid sequence of SEQ ID NO: 8.

In one embodiment, there is provided a polypeptide having endoprotease activity comprising or consisting of a polypeptide having at least 80% sequence identity, such as 85% sequence  
15 identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of SEQ ID NO: 8.

In one embodiment, there is provided a fragment of polypeptide, the fragment having  
20 endoprotease activity, wherein the fragment comprises or consists of a fragment of the amino acid sequence of SEQ ID NO: 1, or a fragment of an amino acid sequence having at least 80% sequence identity, such as 85% sequence identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of SEQ ID NO: 8.

25 In one embodiment, there is provided a polypeptide having endoprotease activity, the polypeptide comprising:

- a) the amino acid sequence of SEQ ID NO: 8;
- b) a variant thereof having at least 80% sequence identity to the amino acid  
30 sequence of SEQ ID NO: 8; or
- c) a fragment of either a) or b).

In one embodiment, there is provided a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid  
35 sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 8.

A polypeptide according to the invention may be provided in a number of states. For example, the polypeptide may be provided as an isolated polypeptide in solution, as a lyophilised solid (e.g., powder), or conjugated (e.g., to other biomolecules or polymers).

5 In one embodiment, the polypeptide is lyophilised.

In one embodiment, the polypeptide is provided in solution.

In one embodiment, the polypeptide is conjugated to a further molecule.

10

In one embodiment, the polypeptide is immobilised on a solid (e.g., a solid support resin) or substantially solid support (e.g., a bead).

#### General features of polypeptides

15 A "polypeptide" is used herein in its broadest sense to refer to a compound of two or more substituent amino acids, amino acid analogues, or other peptidomimetics. The term "polypeptide" therefore includes short peptide sequences as well as longer polypeptides and proteins. The terms "protein", "peptide" and "polypeptide" may be used interchangeably. As used herein, the term "amino acid" refers to either natural and/or unnatural or synthetic  
20 amino acids, including both D or L optical isomers, and amino acid analogues and peptidomimetics. Thus a polypeptide according to the invention may comprise entirely natural or synthetic amino acids, or a mixture of natural and synthetic amino acids.

A polypeptide may be produced by any suitable method, including recombinant or synthetic  
25 methods. For example, the polypeptide may be synthesised directly using standard techniques known in the art, such as Fmoc solid phase chemistry, Boc solid phase chemistry or by solution phase peptide synthesis. Alternatively, a polypeptide may be produced by transforming a cell, such as a bacterial cell, an insect cell, or a mammalian cell, with a nucleic acid molecule or vector that encodes said polypeptide.

30

Expression of polypeptides of the invention is described herein and would further be readily achievable by the person skilled in the art using standard techniques.

A polypeptide may be derivatised or modified to assist with production, isolation or  
35 purification. For example, where a polypeptide of the invention is produced by recombinant expression in a bacterial host cell, the sequence of the polypeptide may include an additional

methionine (M) residue at the N terminus to improve expression. As another example, the polypeptide of the invention may be derivatised or modified by addition of a ligand which is capable of binding directly and specifically to a separation means. Alternatively, the polypeptide may be derivatised or modified by addition of one member of a binding pair and  
5 the separation means comprises a reagent that is derivatised or modified by addition of the other member of a binding pair. Any suitable binding pair can be used.

In some embodiments, the polypeptide of the invention comprises one or more tags, such as polypeptide tags.

10

In one embodiment, the tag is a cleavable tag.

In one embodiment, the tag permit the specific modification of said tag or the polypeptide to which it is attached.

15

In one embodiment, the tag facilitates the identification and/or isolation of the polypeptide.

In one embodiment, the tag is bound by a binding partner. In some embodiments, the binding partner is a polypeptide such as an antibody or other binding protein (e.g., avidin); a  
20 metal (e.g., Ni); and/or an organic small molecule (e.g., biotin).

The polypeptide according to the invention may include any one or more tag selected from the group consisting of: a poly-histidine tag, an Fc tag; a FLAG tag, a Rho1D4 tag, a HA tag, a strep tag, and an Avi tag. However, any suitable tag known to the skilled person may be  
25 utilised in combination with any polypeptide of the invention.

In one embodiment, the tag facilitates or improves the expression and/or properties (e.g., solubility or polymeric state) of the polypeptide.

30 The polypeptide according to the invention may include any one or more tag selected from the group consisting of: an Fc-tag; a SUMO tag; a GST tag; a lysozyme tag; an MBP tag; a GFP tag, a Myc tag, and a thioredoxin tag.

The tag may be included at either the C or N terminus. Further, the one or more tags may be  
35 separated from one another and/or other sequence elements by one or more linker sequence. Any number and combination of tags may be used.

- In a preferred embodiment where the polypeptide for use in the invention is derivatised or modified by addition of one member of a binding pair, the polypeptide is preferably histidine-tagged or biotin-tagged. Typically, the amino acid coding sequence of the histidine or biotin tag is included at the gene level and the polypeptide is expressed recombinantly in *E. coli*. The histidine or biotin tag is typically present at either end of the polypeptide. It may be joined directly to the polypeptide or joined indirectly by any suitable linker sequence, such as 3, 4 or 5 glycine residues, or a mixture of glycine and serine residues, e.g., a 'GGG' linker.
- The histidine tag typically consists of six histidine residues, however the tag can be longer than, such as up to 7, up to 8, up to 9, up to 10 or up to 20 amino acids in length. Alternatively, the His tag can be shorter, for example 5 or fewer, 4 or fewer, 3 or fewer, 2 or fewer or 1 amino acid(s).
- A polypeptide may be provided in a substantially isolated or purified form. That is, isolated from the majority of the other components present in a cellular extract from a cell in which the polypeptide was expressed. By substantially purified, it will be understood that the polypeptide is purified to at least 50%, 60%, 70%, 80% or preferably at least 90% homogeneity. Purity level may be assessed by any suitable means, but typically involves SDS-PAGE analysis of a sample, followed by Coomassie Blue detection. A polypeptide may be mixed with carriers, diluents or preservatives which will not interfere with the intended purpose of the polypeptide and still be regarded as substantially isolated or purified. Where a polypeptide is provided in a composition or combination with an additional active component, such as another polypeptide, each said polypeptide will individually be purified to a high level of homogeneity prior to mixing in an appropriate ratio for the intended purpose of each. For example, two polypeptides may be each be purified to at least 90% homogeneity prior to combining in a 1:1 ratio. Other ratios will be readily discernible by those skilled in the art.
- A polypeptide (or mixture thereof) may be provided in lyophilised form, suitable for reconstitution in aqueous solution prior to use. Lyophilised polypeptides or compositions have improved stability, which facilitate longer storage of the polypeptide. A method of preparing a polypeptide (or mixture thereof) in lyophilised form, comprising freeze-drying said polypeptide (or mixture) in a suitable buffer, such as Tris-buffered saline (TBS), is provided herein. A polypeptide is typically substantially purified prior to freeze-drying. The resulting polypeptide (or mixture) in lyophilised form is also provided. A method of preparing a solution of a polypeptide (or mixture), comprising providing the polypeptide (or mixture) in lyophilised form and reconstituting with a suitable carrier or diluent, such as water, is also provided.

A polypeptide may be immobilised using methods known in the art, for example as described in Datta S *et al.*, Enzyme immobilization: an overview on techniques and support materials, 3 *Biotech*, 3(1):1-9 (2013). For example, the polypeptide may be immobilised by adsorption, 5 covalent binding, affinity immobilization or entrapment. Materials that can be used as supports include but are not limited to for example, natural supports such as agarose, collagen, gelatin, cellulose, pectin, sepharose, inorganic materials such as ceramics, silica, glass, activated carbon or charcoal, or synthetic polymers. For example, the polypeptide may be immobilised on sepharose or agarose, optionally provided as a resin.

10

Amino acid identity may be calculated using any suitable algorithm. For example the BLAST algorithm can be used to calculate identity or align sequences, typically utilising default settings. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).

15

The sequence of a polypeptide of the invention may comprise a variant of the amino acid sequence of any one of SEQ ID NOs: 1 to 11, in which modifications, such as amino acid additions, deletions or substitutions are made relative to the sequence of any of SEQ ID NOs: 1 to 11.

20

Unless otherwise specified, the modifications are preferably conservative amino acid substitutions. Conservative substitutions replace amino acids with other amino acids of similar chemical structure, similar chemical properties or similar side-chain volume. The amino acids introduced may have similar polarity, hydrophilicity, hydrophobicity, basicity, 25 acidity, neutrality or charge to the amino acids they replace. Alternatively, the conservative substitution may introduce another amino acid that is aromatic or aliphatic in the place of a pre-existing aromatic or aliphatic amino acid. Conservative amino acid changes are well-known in the art and may be selected in accordance with the properties of the 20 main amino acids as defined in Table 3 below. Where amino acids have similar polarity, this can 30 be determined by reference to the hydropathy scale for amino acid side chains (see also Table 4).

A sequence of a polypeptide of the invention may comprise a variant of the amino acid sequence of any one of SEQ ID NOs: 1 to 11 in which up to 10, such as up to 20, up to 30, 35 up to 40, up to 50, up to 60, up to 70, up to 80, up to 90, up to 100, up to 150, or up to 200, conservative substitutions are made.

Table 4: Chemical properties of amino acids

| Amino Acid | Chemical properties                                                      | Amino Acid | Chemical properties                                            |
|------------|--------------------------------------------------------------------------|------------|----------------------------------------------------------------|
| Ala (A)    | aliphatic, hydrophobic, neutral, hydrophathy scale value: 1.8            | Met (M)    | hydrophobic, neutral, hydrophathy scale value: 1.9             |
| Cys (C)    | polar, hydrophobic, neutral, hydrophathy scale value: 2.5                | Asn (N)    | polar, hydrophilic, neutral, hydrophathy scale value: -3.5     |
| Asp (D)    | polar, hydrophilic, charged (-), hydrophathy scale value: -3.5           | Pro (P)    | hydrophobic, neutral, hydrophathy scale value: -1.6            |
| Glu (E)    | polar, hydrophilic, charged (-), hydrophathy scale value: -3.5           | Gln (Q)    | polar, hydrophilic, neutral, hydrophathy scale value: -3.5     |
| Phe (F)    | aromatic, hydrophobic, neutral, hydrophathy scale value: 2.8             | Arg (R)    | polar, hydrophilic, charged (+), hydrophathy scale value: -4.5 |
| Gly (G)    | aliphatic, neutral, hydrophathy scale value: -0.4                        | Ser (S)    | polar, hydrophilic, neutral, hydrophathy scale value: -0.8     |
| His (H)    | aromatic, polar, hydrophilic, charged (+), hydrophathy scale value: -3.2 | Thr (T)    | polar, hydrophilic, neutral, hydrophathy scale value: -0.7     |
| Ile (I)    | aliphatic, hydrophobic, neutral, hydrophathy scale value: 4.5            | Val (V)    | aliphatic, hydrophobic, neutral, hydrophathy scale value: 4.2  |
| Lys (K)    | polar, hydrophilic, charged(+), hydrophathy scale value: -3.9            | Trp (W)    | aromatic, hydrophobic, neutral, hydrophathy scale value: -0.9  |
| Leu (L)    | aliphatic, hydrophobic, neutral, hydrophathy scale value: 3.8            | Tyr (Y)    | aromatic, polar, hydrophobic, hydrophathy scale value: -1.3    |

### Polynucleotides

The present invention also relates to polynucleotides that encode polypeptides having IgM specific endoprotease activity.

In one aspect of the invention, there is provided a polynucleotide encoding a polypeptide having IgM specific endoprotease activity.

- 10 The term polynucleotide as used herein is intended to encompass a polymer of nucleotides or nucleotide analogues. Polynucleotides therefore comprise, for example DNA, RNA, or DNA/RNAs.

In one embodiment, the polynucleotide is an RNA.

In one embodiment the polynucleotide is a DNA.

- 5 In one embodiment, the polynucleotide comprises one or more modified nucleotides or nucleotide analogues.

Polynucleotides according to the invention may be naturally occurring, naturally derived, or synthetic.

10

In one embodiment, there is provided a polynucleotide encoding a polypeptide having endoprotease activity comprising or consisting of the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

- 15 In one embodiment, there is provided a polynucleotide encoding a polypeptide having endoprotease activity comprising or consisting of a polypeptide having at least 80% sequence identity, such as 85% sequence identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11..

20

In one embodiment, there is provided a polynucleotide encoding a fragment of polypeptide, the fragment having endoprotease activity, wherein the fragment comprises or consists of a fragment of the amino acid sequence of any one of SEQ ID NOs: 1 to 11., or a fragment of an amino acid sequence having at least 80% sequence identity, such as 85% sequence  
25 identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

- 30 In one embodiment, there is provided a polynucleotide encoding a polypeptide having endoprotease activity, the polypeptide comprising:

- a) the amino acid sequence of any one of SEQ ID NOs: 1 to 11;
- b) a variant thereof having at least 80% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11; or
- c) a fragment of either a) or b).

35

In one embodiment, there is provided a polynucleotide encoding a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of

an amino acid sequence having at least 80% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11.

5 In one embodiment, there is provided a polynucleotide comprising the sequence of any one of SEQ ID NOs: 12 to 22.

In one embodiment, there is provided a polynucleotide consisting of the sequence of any one of SEQ ID NOs: 12 to 22.

10 In one embodiment, there is provided a polynucleotide having at least 60% sequence identity, such as 70% sequence identity, such as 80% sequence identity such as 85% sequence identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to sequence of any of SEQ ID NOs: 12 to 22.

15

In one embodiment, there is provided a polynucleotide comprising:

- a) the polynucleotide sequence of any one of SEQ ID NOs: 12 to 22;
- b) a variant thereof having at least 80% sequence identity to the polynucleotide sequence of any of SEQ ID NOs: 12 to 2; or
- 20 c) a fragment of either a) or b).

In one embodiment, there is provided a polynucleotide encoding a polypeptide having human IgM specific endoprotease activity, wherein the polypeptide comprises or consists of an amino acid sequence having at least 80% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11 and optionally wherein the polynucleotide comprises or consists of a sequence according to any one of SEQ ID NOs: 12 to 22, or a sequence having at least 80% sequence identity to a sequence according to any one of SEQ ID NOs: 12 to 22.

30 The invention also provides for vectors comprising any of the foregoing polynucleotides.

The invention also provides for vectors encoding any of the foregoing polynucleotides.

In one embodiment, the vector is an expression vector.

35

In one embodiment, the vector is a pET21a+ vector.

General features of polynucleotides

The invention therefore provides nucleic acid molecules, i.e., polynucleotides, and vectors that encode a polypeptide of the invention.

- 5 Exemplary polynucleotide molecules encoding polypeptides disclosed herein are provided as SEQ ID NOs: 12 to 22. However, due to the degeneracy of the genetic code it will be understood that any number of sequences may encode identical polypeptides.

- Polynucleotides of the invention may include at the 5' end a codon for the N terminal methionine (ATG) and, prior to the stop codon (TAA) at the 3' end, codons for a Gly-Ser-Gly linker and a 6x His tag, which may optionally be excluded. The optional inclusion of an additional methionine and a tag are discussed in more detail below.
- 10

- The terms "nucleic acid molecule" and "polynucleotide" are used interchangeably herein and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Non-limiting examples of polynucleotides include a gene, a gene fragment, messenger RNA (mRNA), cDNA, recombinant polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers.
- 15

- 20 A polynucleotide of the invention encodes a polypeptide of the invention and may be provided in isolated or substantially isolated form. By substantially isolated, it is meant that there may be substantial, but not total, isolation of the polypeptide from any surrounding medium. The polynucleotides may be mixed with carriers or diluents which will not interfere with their intended use and still be regarded as substantially isolated.
- 25

- A nucleic acid sequence which "encodes" a selected polypeptide is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* or *in vitro* when placed under the control of appropriate regulatory sequences, for example in an expression vector. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. For the purposes of the invention, such nucleic acid sequences can include, but are not limited to, cDNA from viral, prokaryotic or eukaryotic mRNA, genomic sequences from viral or prokaryotic DNA or RNA, and even synthetic DNA sequences. A transcription termination sequence may be located 3' to the coding sequence.
- 30
- 35

Polynucleotides can be synthesised according to methods well known in the art, as described by way of example in Sambrook *et al* (1989, Molecular Cloning - a laboratory manual; Cold Spring Harbor Press). The nucleic acid molecules of the present invention may be provided in the form of an expression cassette which includes control sequences

5 operably linked to the inserted sequence, thus allowing for expression of the polypeptide of the invention *in vivo* (e.g. in prokaryotic or eukaryotic expression systems). These expression cassettes, in turn, are typically provided within vectors (e.g., plasmids or recombinant viral vectors). Such an expression cassette may be administered directly to a host subject. Alternatively, a vector comprising a polynucleotide of the invention may be  
10 administered to a host subject. Preferably the polynucleotide is prepared and/or administered using a genetic vector. A suitable vector may be any vector which is capable of carrying a sufficient amount of genetic information, and allowing expression of a polypeptide of the invention.

15 The present invention thus includes expression vectors that comprise such polynucleotide sequences. Such expression vectors are routinely constructed in the art of molecular biology and may for example involve the use of plasmid DNA and appropriate initiators, promoters, enhancers and other elements, such as for example polyadenylation signals which may be necessary, and which are positioned in the correct orientation, in order to allow for  
20 expression of a peptide of the invention. Other suitable vectors would be apparent to persons skilled in the art. By way of further example in this regard we refer to Sambrook *et al*.

Sequence identity between two or more polynucleotide sequences may be calculated using  
25 any suitable algorithm. For example the BLAST algorithm can be used to calculate identity or align sequences, typically utilising default settings. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).

### 30 Cells

The invention also provides for cells expressing or capable of expressing polypeptides of the invention.

In one embodiment, there is provided a cell comprising a polynucleotide according to the  
35 invention.

In one embodiment, there is provided a cell comprising a vector according to the invention.

In another embodiment, there is provided a cell comprising a polypeptide according to the invention. Such a cell may comprise the polypeptide intracellularly, such as within inclusion  
5 bodies, or solubly within the cytosol; the polypeptide may be transiently present within the cell prior to secretion to the extracellular environment.

Cells typically include prokaryotic cells such as bacterial cells, for example *E. coli*. Such cells may be cultured using routine methods to produce a polypeptide of the invention.

10

In one embodiment the cell is a BL21 (DE3) STAR cell.

Further, cells according to the invention may be eukaryotic cells, such as insect cells or mammalian cells.

15

A cell according to the invention may have been modified, for example by standard molecular biology techniques, to comprise a polynucleotide, vector, or polypeptide according to the invention, and/or modified to express a polypeptide according to the invention.

## 20 **Compositions and combinations**

The polypeptides, polynucleotides, vectors and cells of the invention may be provided in a composition or combination with other compounds.

The invention provides for compositions or combinations comprising the polypeptide of the  
25 invention and a further polypeptide.

By composition, it will be understood that the components of said composition may be mixed; by combination it will be understood that the components of said combination are not necessarily mixed and may be provided separately, e.g., in separate containers, which may  
30 optionally be mixed.

In one embodiment, there is provided a composition or combination comprising a first polypeptide according to the invention, and a second polypeptide, wherein the second polypeptide has IgG specific endoprotease activity.

35

- In one embodiment, the second polypeptide having IgG specific endoprotease activity comprises or consists of a polypeptide having a sequence as set forth in any one of SEQ ID NOs: 23 to 25.
- 5 In one embodiment, there is provided a composition or combination comprising a first polypeptide according to the invention, and a second polypeptide, wherein the second polypeptide has IgG specific endoprotease activity and comprises a sequence having at least 85% sequence identity, such as at least 90%, at least 95%, at least 98%, or at least 99% to the sequence as set forth in SEQ ID NO: 23, or a fragment thereof.
- 10 In particular, it is envisaged that the polypeptide of the invention may be used in conjunction with a second polypeptide having IgG specific endoprotease activity, wherein the second polypeptide comprises or consists of a sequence that lacks the C-terminal poly-histidine tag relative to SEQ ID NOs: 23 to 25.
- 15 In one embodiment, the second polypeptide having IgG specific endoprotease activity is IdeS (FabRICATOR), FabALACTICA, and/or Xork.

Table 5: Exemplary IgG specific proteases

| Sequence identifier | Name(s)           | Polypeptide sequence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|---------------------|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SEQ ID NO: 23       | IdeS (FabRICATOR) | MASMTGGQQMGRDPLADSFSANQEIRYSEVTPYHVT<br>VWTKGVTTPANFTQGEDVFHAPYVANQGWDITKTFNG<br>KDDLCCGAATAGNMLHWWFDQNKDQIKRYLEEHPEKQ<br>KINFNGEQMFDVKEAIDTKNHQLDSKLFYFKEKAFPYL<br>STKHLGVFPDHVIDMFINGYRLSLTNHGPTPVKEGSKDP<br>RGGIFDAVFTRGDQSKLLTSRHDFKEKNLKEISDLIKKEL<br>TEGKALGLSHTYANVRINHVINLWGADFDNNGNLKAIYV<br>TDSDSNASIGMKKYFVGVSAGKVAISAKEIKEDNIGAQV<br>LGLFTLSTGQDSWNQTNLSLEHHHHHH                                                                                                                                                                                                 |
| SEQ ID NO: 24       | FabALACTICA       | MNQNNIQETNLVEKNSEDKFIQELNRYKTEIPNFKGFNV<br>WILGDKGYKYNLINLEEIKNIQATLKERNEEYVFVKLNG<br>KIAHDTTVFLMNKKHKLLKNIEEFKTITQKRLTERGKFPY<br>DTVHSTFEIKDENFIMERLKSSGLSMGKPVDMGVNGIPI<br>YTKTLSIDNKFAFENNSKDSSYSSNINISEDKIKENDQKIL<br>DLIVKSGANNQNLDEEKVIAFTKYIGEITNYDNEAYRAR<br>NVDTEYYRASDLFSVTERKLAMCVGYSVTAARAFNIMGI<br>PSYVVSGKSPQGISHAAVRAYYNRSWHIIDITASTYWKN<br>GNYKTTYSDFIKEYCIDGYDVYDPAKTNNRFVKYMESN<br>EAFENWIHNNGSKSMLFINESAALKDKKPKDDFVPVTEK<br>EKNELIDKYKKLLSQIPENTQNPGEKNIRDYLNKEYEELK<br>KDNLFEEHEAEFKESLNLNESFYLQLKKEKKPSDNLKK<br>EEKPRENSVKERETPAENNDFVSVTEKNLIDKYKELLS |

|                  |      |                                                                                                                                                                                                                                                                                                                                                               |
|------------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                  |      | KIPENTQNPGEKNIRNYLEKEYEELLQKDKLFKHEYTEFT<br>KSLNLNETFYSQLKEGEMKLSENPEKGETNTNGSGLEH<br>HHHHH                                                                                                                                                                                                                                                                   |
| SEQ ID NO:<br>25 | Xork | MSKRKLLKKIEKKDTSSVLTQKKQTKTLWADGVQVDDK<br>DFKPSTENFGTNYLAAEYGIGKGYDINKKFDGTDDDL<br>SGVVAANQLHWWLDRNKDYIEKYRQQSKDNGVTIGNTD<br>IFELNKLHDEDQSNFFDFIKKSFGNKFLQPERLLNMYING<br>YGYLTSQDKAKTSQPSPSKLNFFQKVFKNNLLTDKTPIN<br>DIDEFSSQTKNALQNHKVLAVSFASIKNRGLGHVVTVWG<br>ADFDENGKVVALYVTDSDDRSKNIGNAKLGMKKLRIEVS<br>AQDSSTIKLTGFEDKNSSGSLRHLVSLSTGEQIWKKYFE<br>ETHHHHHH |

The invention also encompasses compositions comprising polypeptides, polynucleotides, vectors and/or cells of the invention in admixture with aqueous components suitable for the storage thereof, such as a diluent, salt, and/or buffer.

5

Further provided herein are pharmaceutical compositions comprising polypeptides, polynucleotides, vectors and/or cells of the invention in admixture with a pharmaceutically acceptable diluent, salt, and/or carrier.

10     **Kits**

Provided herein is a kit comprising any polypeptide, polynucleotide, composition, combination, and/or cell according to the invention. In certain embodiments, the kit comprises packaging and/or instructions for use of said kit.

- 15     In one embodiment, there is provided a kit comprising any combination of one or more of:
- a) a polypeptide according to the invention;
  - b) a composition according to the invention;
  - c) a combination according to the invention;
  - d) a polynucleotide according to the invention;
  - 20     e) a vector according to the invention; and/or
  - f) a cell according to the invention.

In embodiments wherein the kit comprises a combination, the first polypeptide and the second polypeptide may be provided separately (i.e., in different containers within the kit or

25     in different kits) or together (i.e., within the same kit and/or within the same container within the kit).

It will be understood that any substituent parts of a kit according to the invention may be used in the methods or uses of the invention as defined herein. Further, a kit according to the invention may comprise directions for use that relate to any one of the methods or uses of the invention.

5

### Methods and uses

The invention also provides for methods and uses relating to the polynucleotides, polypeptides, compositions, combinations, vectors, cells, and kits of the invention.

10 Any of the methods or uses herein may be performed *in vivo*, *in vitro* or *ex vivo*.

There is provided herein, a method of hydrolysing IgM, the method comprising contacting a sample comprising IgM with the polypeptide according to the invention, or the composition or combination according to the invention.

15

The invention also provides a method of detecting the presence of IgM in a sample, the method comprising:

- a) contacting a sample with a polypeptide, composition, or combination of the invention; and
  - 20 b) identifying the IgM specific cleavage products,
- wherein the presence of IgM specific cleavage products is indicative of the presence of IgM in the sample.

The invention also provides a method of detecting the presence of IgM in a sample, the method comprising:

- 25 a) contacting a sample with a polypeptide, composition, or combination of the invention; and
  - b) identifying and isolating the IgM specific cleavage products,
- wherein the presence of IgM specific cleavage products is indicative of the presence of IgM
- 30 in the sample.

In some embodiments a method of the invention comprises an isolation step. The isolation step may comprise the use of gel electrophoresis and/or chromatography.

In some embodiments a method of the invention comprises an identification and/or analysis step. The identification and/or analysis step may comprise the use of gel electrophoresis, chromatography, immunoblotting, ELISA and/or mass spectrometry.

- 5 Therefore, in some embodiments the methods of the invention comprising identifying, isolating, and/or step(s) comprise the use of:
- a) gel electrophoresis;
  - b) immunoblotting;
  - c) chromatography
  - 10 d) ELISA; and/or
  - e) mass spectrometry.

The invention also provides for the use of the polypeptide according to the invention, the composition according to the invention, the combination according to the invention, or the kit  
15 according to the invention, for the degradation IgM in a sample.

The invention also provides for the use of the polypeptide according to the invention, the composition according to the invention, the combination according to the invention, or the kit  
20 according to the invention, for the removal of IgM from a sample

The invention also provides for the use of the polypeptide according to the invention, the composition according to the invention, the combination according to the invention, or the kit  
according to the invention, for the identification of IgM in a sample

25 In one embodiment of the foregoing methods or uses the IgM is human IgM.

In one embodiment of the foregoing methods or uses the sample is not an *in vivo* sample.

In one embodiment of the foregoing methods or uses the sample is an *ex vivo* sample.  
30

In one embodiment, of the foregoing methods or uses the sample is a sample derived from a human subject.

The polypeptides of the invention have demonstrable activity in both simple and complex  
35 media, wherein the substrate may be considered substantially pure (e.g., a purified IgM sample) or substantially impure (e.g., a crude or complex sample).

In one embodiment, the IgM in the sample is substantially pure.

In one embodiment, the IgM in the sample is substantially impure.

- 5 In one embodiment, the sample comprises or consists of a complex medium, such as a complex biological medium.

In one embodiment, the sample comprises or consists of whole blood, a blood derived fraction, serum, and/or plasma.

10

In preferred embodiments, the sample is a human sample.

In some embodiments the sample is an *ex vivo* sample.

- 15 In some embodiments the sample is a sample obtained during dialysis of a subject.

In some embodiments the sample is a sample obtained from a human subject, wherein the subject no recognised disease or disorder. In some embodiments the sample is a sample obtained from a human subject, wherein the subject no recognised disease or disorder associated with IgM.

20

In some embodiments the sample is a sample obtained from a human subject, wherein the subject has:

- 25 a) an autoimmune disease or disorder;  
b) cancer;  
c) suspected or known infection with a microorganism.

In some embodiments, the microorganism is a bacterium, a virus, or a parasite.

In some embodiments, the microorganism is a pathogenic organism.

- 30 In some embodiments, the foregoing method comprises a step of identifying and/or isolating the IgM specific cleavage products.

In further embodiments, the method additionally comprises contacting the sample with a polypeptide having IgG specific endoprotease activity.

35

In some embodiments, the methods or uses of the invention may be for the purpose of degradation or removal of IgM from a sample.

5 In some embodiments, the methods or uses of the invention may be for the purpose of a diagnostic assay.

In some embodiments, the methods or uses of the invention may be for the purpose of a diagnostic assay, wherein the assay is an anti-drug antibody (ADA) assay.

10 Anti-drug assays are utilised to identify the presence of antibodies directed to or elicited against a drug of interest. Often, these assays are conducted using bridging ELISAs, which do not distinguish between antibody isotypes. A polypeptide according to the present invention, which can degrade IgM, may be used to remove or inactivate IgM in a sample and therefore facilitate increased resolution in ADA assays, or other similar assays that comprise  
15 IgM within a mix of components, by determining the contribution of IgM to any observed activity.

In some embodiments, the methods or uses of the invention may be for the purpose of a diagnostic assay quality control, such as quality control of recombinant or non-recombinant  
20 monoclonal and/or polyclonal IgM or IgM-derived molecules.

For development of IgM, such as IgM for therapeutic purposes, diagnostic purposes, or IgM-like drugs, it is imperative to perform extensive quality control in order to validate the drug and verify, for example, its sequence and/or the presence of post-translational modifications.  
25 Often, such quality control involves the use of trypsin or similar “bottom-up” approaches, which lack specificity and are therefore can be improved. A polypeptide IgM protease hydrolysing the hinge region of IgM could facilitate such analysis.

30 In some embodiments, the methods or uses of the invention may be for the purpose of gene therapy.

IgM has been demonstrated to significantly (negatively) affect gene transfer by adenovirus type 5, which is one of the most commonly used adenovirus vectors. IgM reduces the transduction of the vector in the liver. As such, degradation, inactivation, or removal of IgM  
35 using the polypeptides of the invention may aid gene therapy applications, in particular those utilising adenovirus type 5 vectors.

In one embodiment, there is provided a method or use according to the invention, wherein the method or use is for the removal of IgM during gene therapy. "During" may refer to the pre-treatment, concomitant treatment, or post-treatment of a subject or sample with a polypeptide of the invention.

5

In some embodiments, the methods or uses of the invention may be for the treatment or prevention of cancer.

10 In some embodiments, the methods or uses of the invention may be for the treatment or prevention of diffuse large B-cell lymphoma.

Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of non-Hodgkin lymphoma (NHL) and affects several thousand individuals on a yearly basis, with poor prognosis if left untreated. Even with treatment, the overall five-year survivability is low. The lymphoma B-cells most often have IgM as their B-cell receptor (BCR), and the removal or degradation of this receptor using a polypeptide of the invention may aid in the treatment of the cancer, e.g., due to the induction of apoptosis and the clearance of the lymphoma B-cells.

20 In some embodiments, the methods or uses of the invention may be for the treatment or prevention of an autoimmune disease or disorder.

In some embodiments, the methods or uses of the invention may be for the treatment or prevention of Systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and/or Sjögren's syndrome.

Several autoimmune diseases, including SLE, RA, and Sjögren's syndrome, are characterised by high levels of Rheumatoid factor (RF), which is often used as a prognostic biomarker for the diseases. High levels of RF are associated with worse outcomes and more severe pathogenesis. RF is typically an IgM-based antibody directed towards self IgG, which stimulates the formation of immune complexes that contribute to the disease, e.g., by causing chronic inflammation and joint destruction. The degradation or removal of IgM in such diseases using a polypeptide of the invention may therefore help in the treatment or prevention of these diseases.

35

In one aspect, there is provided a polypeptide, combination, composition, or pharmaceutical composition according to the invention, for use in the manufacture of a medicament.

In one aspect, there is provided a polypeptide, combination, composition, or pharmaceutical composition according to the invention, for use in therapy.

- 5 In one aspect, there is provided a polypeptide, combination, composition, or pharmaceutical composition according to the invention, for use in the treatment or prevention of a disease or disorder associated with elevated IgM.

- 10 In one embodiment, there is provided a polypeptide, combination, composition, or pharmaceutical composition according to the invention, for use in the treatment or prevention of cancer.

- 15 In one embodiment, there is provided a polypeptide, combination, composition, or pharmaceutical composition according to the invention, for use in the treatment or prevention of diffuse large B-cell lymphoma.

- 20 In one embodiment, there is provided a polypeptide, combination, composition, or pharmaceutical composition according to the invention, for use in the treatment or prevention of an autoimmune disease or disorder.

- In one embodiment, there is provided a polypeptide, combination, composition, or pharmaceutical composition according to the invention, for use in the treatment or prevention of Systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and/or Sjögren's syndrome.

- 25 In one embodiment, there is provided a polypeptide, combination, composition, or pharmaceutical composition according to the invention, for use in the treatment or prevention of IgM nephropathy.

**ASPECTS OF THE INVENTION**

The invention will be described by way of the following numbered aspects.

Aspect 1. A polypeptide having endoprotease activity comprising:

- 5                   a) the amino acid sequence of SEQ ID NO: 1;  
                    b) a variant thereof having at least 80% sequence identity to the amino acid  
                    sequence of SEQ ID NO: 1; or  
                    c) a fragment of either a) or b).

10   Aspect 2. A polypeptide having endoprotease activity comprising:

- a) the amino acid sequence of SEQ ID NO: 8;  
                    b) a variant thereof having at least 80% sequence identity to the amino acid  
                    sequence of SEQ ID NO: 8; or  
                    c) a fragment of either a) or b).

15

Aspect 3. the polypeptide according to aspect 1 or aspect 2 wherein the variant of said polypeptide has at least 85% sequence identity, at least 90% sequence identity, at least 91% sequence identity, at least 92% sequence identity, at least 93% sequence identity, at least 94% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least 97% sequence identity, at least 98% sequence identity, or at least 99% sequence identity to the amino acid sequence of SEQ ID NO: 1 or SEQ ID NO: 8.

20

Aspect 4. The polypeptide according to aspect 1 or aspect 3, wherein the polypeptide comprises or consists of:

- 25                   a) the amino acid sequence of SEQ ID NO: 2;  
                    b) a variant thereof having at least 80% sequence identity to the amino acid  
                    sequence of SEQ ID NO: 2; or  
                    c) a fragment of either a) or b).

30   Aspect 5. The polypeptide according to aspect 2 or aspect 3, wherein the polypeptide comprises or consists of:

- a) the amino acid sequence of SEQ ID NO: 9;  
                    b) a variant thereof having at least 80% sequence identity to the amino acid  
                    sequence of SEQ ID NO: 9; or  
35                   c) a fragment of either a) or b).

Aspect 6. The polypeptide according to any of aspects 1 to 5, wherein the fragment of said polypeptide is at least 100 amino acids in length, at least 150 amino acids in length, at least 200 amino acids in length, at least 250 amino acids in length, at least 300 amino acids in length, at least 350 amino acids in length, at least 400 amino acids in length, at least 450 amino acids in length, at least 500 amino acids in length, at least 550 amino acids in length, at least 600 amino acids in length, or at least 650 amino acids in length.

Aspect 7. The polypeptide according to any of aspects 1 to 6, wherein the polypeptide includes:

- a) an additional methionine at the N terminus; and/or
- b) a His tag.

Aspect 8. The polypeptide according to aspect 7, wherein the His tag is at the C terminus.

Aspect 9. The polypeptide according to aspect 7, wherein the His tag is at the N terminus.

Aspect 10. The polypeptide according any of aspects 1 to 8, wherein the polypeptide comprises or consists of a sequence selected from the group consisting of SEQ ID NOs: 3, 5, 6, 7, and 10.

Aspect 11. The polypeptide according any of aspects 1 to 7, and 9, wherein the polypeptide comprises or consists of a sequence selected from the group consisting of SEQ ID NOs: 4 and 11.

Aspect 12. The polypeptide according any of aspects 1 to 8, and 10, wherein the polypeptide comprises or consists of SEQ ID NO: 7.

Aspect 13. The polypeptide according to any of aspects 1 to 12, wherein said endoprotease activity is IgM specific endoprotease activity.

Aspect 14. The polypeptide according to any of aspects 1 to 13, wherein the polypeptide cleaves human IgM.

Aspect 15. The polypeptide according to any of aspects 1 to 14, wherein the polypeptide is provided in solution, lyophilised, or immobilised.

Aspect 16. A composition or combination comprising a first polypeptide according to any of aspects 1 to 15, and a second polypeptide, wherein the second polypeptide has IgG specific endoprotease activity.

5

Aspect 17. A polynucleotide encoding a polypeptide according to any of aspects 1 to 15.

Aspect 18. A vector encoding or comprising a polynucleotide according to aspect 17.

10 Aspect 19. A cell comprising the polypeptide of any of aspects 1 to 15, the polynucleotide of aspect 17, or the vector of aspect 18.

Aspect 20. The cell according to aspect 20, wherein the cell is a bacterial cell.

15 Aspect 21. A kit comprising:

- a) the polypeptide according to any of aspects 1 to 15;
- b) the composition or combination according to aspect 16, wherein the first polypeptide and the second polypeptide are provided separately or together;
- c) the polynucleotide according to aspect 17;
- 20 d) the vector according to aspect 19; and/or
- e) the cell according to aspect 20.

Aspect 22. A method of hydrolysing IgM, the method comprising contacting a sample comprising IgM with the polypeptide according to any of aspects 1 to 15, or the composition  
25 or combination of aspect 16.

Aspect 23. The method according to aspect 22, wherein the IgM is human IgM.

30 Aspect 24. The method according to aspect 22, wherein the IgM is non-human primate IgM, preferably monkeys of the genus *Macaca*, such as *Macaca mulatta* and/or *Macaca fascicularis*, and/or the genus *Papio*.

Aspect 25. The method according to any of aspects 22 to 24, wherein the sample is:

- a) substantially pure; or
- 35 b) substantially impure.

Aspect 26. The method according to any of aspects 22 to 25, wherein the sample comprises a complex biological medium.

5 Aspect 27. The method according to any of aspects 22 to 26, wherein the sample comprises blood, serum, and/or plasma.

Aspect 28. The method according to any of aspects 22 to 27, wherein the sample comprises human blood, human serum, and/or human plasma.

10 Aspect 29. The method according to any of aspects 22 to 28, wherein the method comprises a step of identifying and/or isolating the IgM specific cleavage products.

Aspect 30. The method according to any of aspects 22 to 29, wherein the method additionally comprises contacting the sample with a polypeptide having IgG specific  
15 endoprotease activity, optionally wherein the polypeptide having IgG specific endoprotease activity comprises or consists of a polypeptide having a sequence as set forth in any one of SEQ ID NOs: 23 to 25.

Aspect 31. A method of detecting the presence of IgM in a sample, the method comprising:  
20 a) contacting a sample with the polypeptide according to any of aspects 1 to 15 or the composition or combination according to aspect 16; and  
b) identifying and/or isolating the IgM specific cleavage products,  
wherein the presence of IgM specific cleavage products is indicative of the presence of IgM  
in the sample.

25 Aspect 32. The method according to any of aspects 22 to 31, wherein the identifying and isolating steps comprise the use of:

- 30 a) gel electrophoresis;  
b) immunoblotting;  
c) chromatography;  
d) ELISA and/or  
e) mass spectrometry.

Aspect 33. Use of the polypeptide according to any of aspects 1 to 15, the composition or  
35 combination according to aspect 16, or the kit according to aspect 21, for the degradation, removal, or identification of IgM in a sample.

Aspect 34. The method according to any of aspects 22 to 32, or the use according to aspect 33, wherein the sample is an *in vitro*, *in vivo*, or *ex vivo* sample.

- Aspect 35. The method according to any of aspects 22 to 32, or the use according to aspect 33 or aspect 34, wherein the method or use is:
- a) for quality control (QC) processes; and/or
  - b) for diagnostic assays.

- Aspect 36. The method according to any of aspects 22 to 32 and 35, or the use according to aspect 33 to 35, wherein the method or use is:
- a) for the degradation or removal of IgM during dialysis;
  - b) for quality control of recombinant or non-recombinant monoclonal and/or polyclonal IgM or IgM-derived molecules; and/or
  - c) for anti-drug antibody (ADA) assays;

Aspect 37. The method or use according to any of aspects 22 to 36, wherein the method or use is non-therapeutic.

Aspect 38. The method or use according to any of aspects 22 to 36, wherein the method or use is therapeutic.

- Aspect 39. The method according to any of aspects 22 to 32, 35 and 38, or the use according to aspects 33 to 36 and 38, wherein the method or use is for the treatment or prevention of:
- a) kidney disease or dysfunction;
  - b) cancer; and/or
  - c) autoimmune diseases or disorders.

- Aspect 40. The method or use according to aspect 39, wherein the method or use is for the treatment or prevention of:
- a) diffuse large B-cell lymphoma;
  - b) systemic lupus erythematosus (SLE);
  - c) rheumatoid arthritis (RA);
  - d) Sjögren's syndrome; and/or
  - e) IgM nephropathy.

Aspect 41. A polypeptide according to any of aspects 1 to 15, composition or combination according to aspect 16, or kit according to aspect 21, for use in therapy.

5 Aspect 42. A polypeptide according to any of aspects 1 to 15, composition or combination according to aspect 16, or kit according to aspect 21, for use in the treatment or prevention of a disease or disorder, wherein the disease or disorder is:

- a) kidney disease or dysfunction;
- b) cancer; and/or
- c) an autoimmune disease or disorder.

10

Aspect 43. The polypeptide, composition, combination, or kit according to aspect 42, wherein the disease or disorder is:

- a) diffuse large B-cell lymphoma;
- b) systemic lupus erythematosus (SLE);
- 15 c) rheumatoid arthritis (RA);
- d) Sjögren's syndrome; and/or
- e) IgM nephropathy.

20 Aspect 44. A polypeptide according to any of aspects 1 to 15, composition or combination according to aspect 16, or kit according to aspect 21, for use in the manufacture of a medicament.

25 Aspect 45. The polypeptide, composition, combination, or kit for use according to aspect 44, wherein the medicament is for the treatment or prevention of a disease or disorder, wherein the disease or disorder is:

- a) kidney disease or dysfunction;
- b) cancer; and/or
- c) an autoimmune disease or disorder.

30 Aspect 46. The polypeptide, composition, combination, or kit for use according to aspect 44 or aspect 45, wherein the medicament is for the treatment or prevention of a disease or disorder, wherein the disease or disorder is:

- a) diffuse large B-cell lymphoma;
- b) systemic lupus erythematosus (SLE);
- 35 c) rheumatoid arthritis (RA);
- d) Sjögren's syndrome; and/or

e) IgM nephropathy.

The invention will now be further described by way of Examples, which are meant to serve to assist one of ordinary skill in the art in carrying out the invention and are not intended in any way to limit the scope of the invention.

EXAMPLES

**Example 1: Identification, isolation and purification of an IgM specific protease from *Lachnoanaerobaculum***

- 10 Identification of potential proteases in *Lachnoanaerobaculum*
- The present inventors utilized combined bioinformatic and biochemical analysis to identify homologues of the only known IgM specific endoprotease, IdeSSuis (Seele J., et al., *J Bacteriol* 2013, 195(5)).
- 15 Whilst none of the known isolates of *Lachnoanaerobaculum* have been shown to possess proteolytic or specific immunoglobulin-modifying activity, the present inventors surprisingly found genes containing IdeS-like domains that exhibited low levels of sequence similarity to the known IgM specific endoprotease IdeSSuis.
- 20 The identification of IdeS-like domain containing genes was surprising for two reasons: (1) these domains have previously been primarily identified in *Streptococcus*, and (2) the presence of IdeS-like domain containing enzymes is associated with pathogenic bacteria, not commensal organisms that form part of the normal human flora.

Synthesis of candidate proteins

- 25 A number of constructs were designed, synthesized, and used to express the desired protein in a recombinant host cell.

Constructs were designed and synthesized in accordance with Table 5.

Table 5: Design and synthesis of polypeptides

| Construct                | Short name | Polypeptide sequence |
|--------------------------|------------|----------------------|
| L. gingivalis D1-C-6His  | LG-FL      | SEQ ID NO: 9         |
| L. gingivalis D1-N-10His | LG-FL10    | SEQ ID NO: 11        |
| L. umeaense D1-C-6His    | LU-FL      | SEQ ID NO: 3         |
| L. umeaense D1-N-10His   | LU-FL10    | SEQ ID NO: 4         |

|                                     |         |              |
|-------------------------------------|---------|--------------|
| L. umeaense Suis-C-6His             | LU-S    | SEQ ID NO: 5 |
| L. umeaense Suis $\Delta$ 38-C-6His | LU-d38S | SEQ ID NO: 6 |
| L. umeaense MAC-1-C-6His            | LU-M1   | SEQ ID NO: 7 |

A full-length construct of the IdeS/Mac-1-containing gene from both *L. gingivalis* and *L. umeaense* was expressed, with either an N- or C-terminal His-tag (*L. gingivalis* D1-N-10H, *L. umeaense*-D1-N-10H, *L. gingivalis* D1-C-6His, and *L. umeaense*-D1-C-6His, respectively).

- 5 The full-length constructs exhibited yields of less than 5mg/g, and the resultant polypeptides were of relatively low purity and fragmented.

The construct from *L. umeaense* was selected to be further engineered to only express certain fragments of the full length protein. The *L. umeaense* Suis (LU-S) construct expressed well and generated high yields upon optimization of the purification process, but generated a protein that was prone to fragmentation under the conditions tested. A fragment that was 38 amino acid shorter (Suis $\Delta$ 38; LU-d38S) was more stable but yields were low, under the conditions tested.

- 15 The constructs were further engineered and optimized until a construct, (*L. umeaense* MAC-1; further called LU-M1) was found that expressed with high yield (>20 mg/g), high purity, and homogeneity (Figure 1). Expression of LU-M1 was scalable, when expressed in bioreactors, the yield of LU-M1 was approximately 10 mg/g, with the cultures reaching high-density (220 g cells per 600 mL batch).

20

LU-M1 shares little sequence identity with a number of known streptococcal virulence factors (Table 6). However, the construct displays much higher levels of sequence identity with the *L. gingivalis* IdeS/Mac-1-containing protein.

- 25 Table 6: Sequence identity between LU-M1 and other bacterial proteases

|                                 | EndoS | IdeS | IdeZ | Xork | IdeSSuis | <i>L. gingivalis</i> |
|---------------------------------|-------|------|------|------|----------|----------------------|
| <b><i>L. umeaense</i> MAC-1</b> | 17%   | 31%  | 30%  | 28%  | 45%      | 74%                  |

All constructs were codon-optimized for *E. coli* and inserted into a pET21a+ backbone before transformed into BL21 (DE3) STAR cells under ampicillin selection. Clones were cultivated in LB supplemented with ampicillin, and expression of protein was conducted in baffled 2 L flasks and induced at OD<sub>600</sub> = 0.5-0.7 with addition of 1 mM IPTG. After 4 hours of induction at 20-37°C, the cells were harvested by centrifugation and kept at -20°C until

30

further processing. Cells were lysed by sonication, and the lysate cleared from cell debris through centrifugation. The material was purified on a His GraviTRAP, and purity assessed by SDS-PAGE.

- 5 For expression in a bioreactor system, LU-M1 was produced recombinantly using *Escherichia coli* BL21(DE3) STAR. Fermentation was performed using a high-cell density fed-batch strategy, with chemically defined minimal media. Induction, harvesting, and purification steps were analogous to those used for small scale production.

10 **Example 2: Functional characterization of an IgM specific protease from *Lachnoanaerobaculum***

- To investigate the activity of the full-length constructs (LG-FL and LU-FL) the purified proteins were incubated with human and mouse IgG (Figure 2A). Only LG-FL exhibited some activity against human serum IgG, which was further demonstrated to be specifically  
15 against human IgG<sub>2</sub> (Figure 2B).

- Despite homology to known IgG proteases and due limited (LG-FL) or no (LU-FL) activity displayed against human IgG, purified LU-FL and LG-FL were incubated with human IgA and IgM. Surprisingly, both LG-FL and LU-FL were active against IgM; and no IgA-specific  
20 activity was observed (Figure 2C).

The IgM-specific activity observed in the full-length constructs was retained by the shorter engineered fragment LU-M1 (Figure 2D).

- 25 Furthermore, LU-M1 was shown not to have activity against the broad protease substrate casein (Figure 2E) or IgM from rodent species (mouse, rabbit, rat). However, LU-M1 displayed some activity against IgM purified from monkey (Figure 2F). Purified IgM from monkeys was obtained from Rockland (017-0107) and is a combination of rhesus, cynomolgus monkey and baboon IgM.

- 30 General protease activity was determined using the EnzCheck Protease Assay Kit, according to manufacturer's instructions. As a positive control the general protease SpeB (FabULOUS) was used, and as a negative control the IgG-specific protease IdeS (FabRICATOR). All reactions were measured in a fluorescence microplate reader after 1 h  
35 and 2 h incubation at 37°C.

Further, for the assessment of species-specificity, IgM from different species were incubated at 37°C for 1 hour (1:40, w:w, enzyme:substrate) with LU-M1 prior to SDS-PAGE analysis.

5 Myeloma IgM was deglycosylated with PNGaseF, and incubated with LU-M1 to investigate the site of hydrolysis (Figure 3). Only the heavy chain was hydrolyzed, while the light chain and the J-chain remained intact. The intact mass sum of the deglycosylated heavy chain fragment (63140.6167 Da) aligned well with the measured mass of the reduced IgM heavy chain (63140.9971 Da). Upon addition of LU-M1 the heavy chain was fragmented into m/z 37499.6142 and m/z 25659.0200. The C-terminal fragment (25659.0200 Da) could be  
10 assigned to the theoretical mass value of amino acids 221-453 (25657.6646 Da) of human IgM constant region (UniProt accession: P01871).

IgM myeloma (Sigma-Aldrich) was incubated with and without addition of purified recombinant LU-M1 at 37°C for 1 h using an enzyme to substrate ratio of 1:40 (w:w) in PBS  
15 buffer. The mixtures were denatured and reduced in 50 mM DTT (GBiosciences), 0.5% (v/v) laurylsarcosine at 90°C for 5 min and cooled to room temperature before addition of PNGaseF (Genovis) at 2 U/μg. The N-deglycosylation reactions were allowed to proceed for 2 h at 37°C before the samples were dialysed against TBS using a Slide-A-Lyzer MINI device (ThermoFisher Scientific). The hydrolyzed material was denatured in 4 M GnHCl  
20 (Sigma-Aldrich) and 100 mM DTT at 37°C for 30 min. The digestion products were separated on a reversed phase C4 column (Acquity premier BEHC4, 450Å, 2.7 μm 2.1 x 100 mm, Waters) using an acetonitrile gradient and analyzed by ESI-QTOF-MS (Bruker Impact II). Deconvolution was preformed using Bruker Compass DataAnalysis version 5.2 and the MaxEnt algorithm.

25 Digestion takes place below the CH2 region of human IgM (...VPDQDT / AIRVFA...), similar to the porcine IgM protease IdeSSuis. The native reaction leaves F(ab')<sub>2</sub> (VH-CH1-CH2) and a pentameric Fc (CH3-CH4), due to the inter-monomer disulphide bonds between cysteine 413 residues in CH3 (Figure 4).

30

### Example 3: Optimization and analysis of protease activity

In order to further assess the activity of the LU-M1, enzymatic activity of the polypeptide was assessed in respect of pH, salt (NaCl), divalent cations, temperature, and susceptibility/resistance to protease inhibitors.

35

LU-M1 is capable of hydrolyzing IgM under a wide range of pH conditions (5.5-9.0; Figure 5A) without significant reduction in activity.

Enzymatic activity appears negatively impacted by increasing NaCl concentration, a 12.5%  
5 reduction in activity in the presence of 150 mM NaCl, a 40% reduction at 500 mM NaCl, and  
a 75% reduction at 1000 mM NaCl, was observed (Figure 5B).

LU-M1 is highly active in the absence of any added divalent cations (Figure 5C). Rather, the  
activity appears slightly reduced by addition of  $Mg^{2+}$  and  $Ca^{2+}$  ions. Aggregation was  
10 observed in the presence of  $Zn^{2+}$ . Further, the addition of EDTA appears to cause a  
concentration-dependent reduction in enzymatic activity (30-50% reduction, depending on  
concentration).

Enzymatic activity is enhanced by increasing the temperature; nevertheless, LU-M1 retains  
15 some activity at 4°C (35% hydrolysis). Although slower than at 37°C, the enzyme is capable  
of efficiently hydrolyzing IgM at room temperature (Figure 1D).

Next, the activity of LU-M1 was assessed in the presence of a panel of protease inhibitors  
(Figure 5E). Serum IgM was pre-incubated (15 min) with a set of protease inhibitors (G-  
20 Biosciences) before addition of LU-M1 (1:50 w:w LU-M1:IgM), and the mixture was  
incubated for 30 min at 37°C before analysis by SDS-PAGE.

Based on the inhibitory capacity of this panel of protease inhibitors, LU-M1 would be  
classified as a cysteine protease; activity is completely inhibited by the presence of  
25 iodoacetamide, and strongly inhibited by the presence of AEBSF and chymostatin. None of  
the other evaluated protease inhibitors had a significant impact on enzyme activity.

LU-M1 was incubated with IgM (IgM myeloma (20 µg)) at different ratios (1:20, 1:50, and  
1:100; LU-M1 to IgM (w:w)) and incubated for 5 to 60 minutes to investigate the efficiency of  
30 hydrolysis. At low enzyme:substrate ratios (i.e., high relative enzyme concentration), almost  
all IgM was hydrolyzed within 5 minutes; at 1:100 all IgM was hydrolysed within 60 minutes  
(Figure 6A, B). Similarly, a higher concentration of substrate results in a faster hydrolysis  
(Figure 6C).

35 Unless otherwise stated, for all activities related to optimal conditions, an enzyme:substrate  
ratio of 1:100 was used, with an incubation of 30 min at 37°C.

Next, the ability of the enzyme to function in complex media was assessed. Undiluted serum was incubated with the IgG-specific protease IdeS, the IgM-specific protease LU-M1 as defined herein, or a combination thereof, and the resultant mixtures analysed using SDS-PAGE and western blotting.

5

Human heat-inactivated serum (Sigma, H2667) was incubated with IdeS (1:50, w:w enzyme:IgG) and LU-M1 (1:40, w:w, enzyme:IgM) for 30 min at 37°C. Material was analyzed by SDS-PAGE and western blotting. Anti-human IgM ALP (Mabtech, 3880-9A-1000; 1:1000) and anti-human IgG ALP (Mabtech, 3310-1-1000; 1:2000) were used as primary antibodies, and the western breeze kit (Invitrogen) for blocking, washing, and chromogenic substrate detection.

10

Both IdeS and LU-M1 were found to be highly specific, targeting only IgG and IgM, respectively in the sample. Furthermore, it was observed that both enzymes can operate in the same sample without any significant effect on the function of one another (Figure 7).

15

These data evidence the ability of LU-M1 to catalyze the specific hydrolysis of IgM in substantially impure sample such as complex biological samples like serum.

#### 20 **Example 4: Assessing the ability to manipulate LU-M1**

The ability to alter the state of LU-M1 to facilitate ease of handling LU-M1 in a variety of circumstances was assessed.

25

Firstly, LU-M1 was lyophilized at a variety of concentrations and volumes. All conditions resulted in highly active material, and thus indicate that LU-M1 is stable in a lyophilized format and can be reconstituted to yield an active enzyme (Figure 8).

30

Lyophilization was performed in a Cool-Safe 10-4 Pro (ScanVac). Briefly, aluminum blocks for vials were pre-cooled at -80°C before samples (25-100 µL, 0.5-2 mg/mL) were resuspended in TBS were added to the block and incubated overnight at -80°C. Samples were lyophilized for 24 hours, and pellets were stored at -20°C.

35

Activity of lyophilized material was evaluated by reconstitution of material in milliQ water to 40 U/µL (1 U = 0.02 µg). Reconstituted enzyme was incubated with 1 U/µg IgM for 1 h at 37°C in PBS before SDS-PAGE analysis (Figure 8).

Secondly, LU-M1 was immobilized according to manufacturer's instructions (NHS activated Sepharose, Cytiva), at working concentrations of 20-30 mg/mL. Activity of the immobilized protein was assessed by mixing 50  $\mu$ L resin with 100  $\mu$ g IgM (100  $\mu$ L) and incubating with end-over-end agitation at room temperature for 15-30 minutes. LU-M1 was readily  
5 immobilized at high concentrations (20-30 mg/mL), with good coupling efficiencies (92-94%) and retained activity (Figure 8C).

These data indicate that LU-M1 is readily lyophilized/immobilized and retains functionality after being subject to these processes, thus evidencing the broad applicability of the enzyme  
10 to a number of scenarios and indicating it's suitability for storage and/or transport in forms other than as an isolated aqueous protein.

#### **Example 5: Assessing reactivity of IVIG against LU-M1**

The presence of LU-M1 specific human IgG was assessed using SDS-PAGE analysis and  
15 western blotting (fFigure 9).

0.5  $\mu$ g each of IdeS, Xork, and LU-M1 were separated by SDS-PAGE and either or transferred to a nitrocellulose membrane for western blotting. The membrane was blocked in a casein solution before addition of the polyclonal primary antibody mixture, IVIG at 5  $\mu$ g/mL.  
20 Primary antibody was incubated for 2 hours at room temperature. The membrane was washed, and a secondary antibody (alkaline phosphatase conjugated  $\alpha$ -human IgG HC, 1:2000) was added and incubated for 1 hour at room temperature. The membrane was thoroughly washed before addition of the chromogenic substrate BCIP/NBT.

25 The western blot data indicate that little to no anti-LU-M1 antibodies exist in IVIG. The presence of none, or low levels of anti-enzyme antibodies in human IVIG is advantageous and highly desirable. Specifically, an absence or paucity of anti-enzyme antibodies facilitates human treatment since it will increase the number of patients susceptible to treatment and increase the efficiency of treatment, e.g., as enzyme is not removed or does not cause  
30 adverse reactions.

**CLAIMS**

1. A polypeptide having endoprotease activity comprising:
  - a) the amino acid sequence of SEQ ID NO: 1 or SEQ ID NO: 8;
  - b) a variant thereof having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 1 or SEQ ID NO: 8; or
  - c) a fragment of either a) or b).
2. The polypeptide according to claim 1, wherein the polypeptide comprises or consists of:
  - a) the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 9;
  - b) a variant thereof having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 9; or
  - c) a fragment of either a) or b).
3. The polypeptide according to claim 1 or claim 2, wherein the polypeptide includes:
  - a) an additional methionine at the N terminus; and/or
  - b) a His tag;wherein the His tag is at the C terminus or at the N terminus.
4. The polypeptide according any of claims 1 to 3, wherein the polypeptide comprises or consists of a sequence selected from the group consisting of SEQ ID NOs: 3, 4, 5, 6, 7, 10, and 11.
5. The polypeptide according any of claims 1 to 4, wherein the polypeptide comprises or consists of SEQ ID NO: 7.
6. The polypeptide according to any of claims 1 to 5, wherein said endoprotease activity is IgM specific endoprotease activity.
7. The polypeptide according to any of claims 1 to 6, wherein the polypeptide cleaves human IgM.
8. The polypeptide according to any of claims 1 to 7, wherein the polypeptide is provided in solution, lyophilised, or immobilised.
9. A composition or combination comprising a first polypeptide according to any of claims 1 to 8 and a second polypeptide, wherein the second polypeptide has IgG specific endoprotease activity.

10. A polynucleotide encoding a polypeptide according to any of claims 1 to 8.
11. A vector encoding or comprising a polynucleotide according to claim 10.
12. A cell comprising the polypeptide of any of claims 1 to 8, the polynucleotide of claim 10, or the vector of claim 11.
13. A kit comprising:
  - a) the polypeptide according to any of claims 1 to 8;
  - b) the composition or combination according to claim 9, wherein the first polypeptide and the second polypeptide are provided separately or together;
  - c) the polynucleotide according to claim 10;
  - d) the vector according to claim 11; and/or
  - e) the cell according to claim 12.
14. A method of hydrolysing IgM, the method comprising contacting a sample comprising IgM with the polypeptide according to any of claims 1 to 8, or the composition or combination according to claim 10.
15. The method according to claim 14 wherein the IgM is human IgM.
16. The method according to any of claims 13 to 15, wherein the sample comprises a complex biological medium, preferably wherein the sample comprises blood, serum, and/or plasma, more preferably wherein the sample comprises human blood, human serum, and/or human plasma.
17. The method according to any of claims 13 to 16, wherein the method comprises a step of identifying and/or isolating the IgM specific cleavage products.
18. The method according to any of claims 13 to 17, wherein the method additionally comprises contacting the sample with a polypeptide having IgG specific endoprotease activity.
19. A method of detecting the presence of IgM in a sample, the method comprising:
  - a) contacting a sample with the polypeptide according to any of claims 1 to 8 or the composition or combination according to claim 9; and
  - b) identifying and/or isolating the IgM specific cleavage products,

wherein the presence of IgM specific cleavage products is indicative of the presence of IgM in the sample.

20. Use of the polypeptide according to any of claims 1 to 8, the composition or combination according to claim 9, or the kit according to claim 12, for the degradation, removal, or identification of IgM in a sample.