

ENZYMES FOR GLYCAN ANALYSIS

Field of the Invention

The present invention relates to enzymes and combinations thereof useful for
5 studying glycoproteins, and corresponding methods of use.

Background of the Invention

Glycosylation of proteins plays a pivotal role in many physiological functions in humans, including signalling, transportation, protection from proteolytic activity,
10 adherence, inflammatory response, microbial colonization, etc. Most of the glycan chains attached to proteins, whether O- or N-linked, are decorated with terminal sialic acids. Being the outermost glycan on glycoproteins, their presence/absence is often critical for the downstream effect, including for example the inflammatory potential in immuno-
15 globulins. Sialic acids on proteins are heterogenous, both in terms of presence/absence on a given protein, as well as individual structural modifications. They are also generally negatively charged which complicates mass spectrometry analysis. This makes the study of glycoproteins difficult, as well as reducing the ability of manufacturers to confirm that a glycoprotein batch will function in a homologous matter. To overcome these problems attempts have been made to genetically engineer CHO cells to reduce the complexity of
20 glycans, although this may affect function. Chemical approaches have also been used, but these often damage the proteins. There is a need for an alternative approach to remove sialic acid from glycoproteins. Furthermore, once sialic acids are removed, there is a need for more tools to study the remaining glycan chains, particularly those which are O-linked.

Summary of the Invention

The present invention provides:

A composition comprising a first sialidase which is independently selected from:

- (a) a polypeptide comprising or consisting of an amino acid sequence of SEQ ID NO: 2;
- (b) a polypeptide comprising or consisting of an amino acid sequence which is at least
30 85% identical to the amino acid sequence of SEQ ID NO: 2 or

- (c) a polypeptide comprising or consisting of an amino acid sequence which is a fragment of the sequence of SEQ ID NO: 2 or a fragment of an amino acid which is 85% identical to the amino acid sequence of SEQ ID NO: 2; optionally wherein said first sialidase includes an additional methionine at the N terminus and/or a histidine tag at the C terminus, which tag may be joined to the C terminus by a linker.

The present invention also provides a composition as defined above which further comprises a second sialidase which is independently selected from:

- (d) a polypeptide comprising or consisting of an amino acid sequence of SEQ ID NO: 5;
- (e) a polypeptide comprising or consisting of an amino acid sequence which is at least 85% identical to the amino acid sequence of SEQ ID NO: 5 or
- (f) a polypeptide comprising or consisting of an amino acid sequence which is a fragment of the sequence of SEQ ID NO: 5 or a fragment of an amino acid which is 85% identical to the amino acid sequence of SEQ ID NO: 5.
- optionally wherein said second sialidase includes an additional methionine at the N terminus and/or a histidine tag at the C terminus, which tag may be joined to the C terminus by a linker.

The composition may additionally comprises a glycosidase and/or a protease, which is optionally present in highly purified or isolated form.

- The present invention also provides a method for modifying a glycoprotein comprising contacting a sample containing the glycoprotein with a as defined above optionally wherein the resulting products are analysed.

Brief Description of the Figures

- Figure 1:** Expression and purification of *Akkermansia* sialidases. All four sialidases, herein represented by Am0707 and Am1757, expressed well and were able to be purified to high homogeneity. Values (mg/mL) indicate the concentration obtained after His-purification.

- Figure 2:** Sialidase activity upon different sialic acid bonds. The two sialidase products (Am1757, and Mix (Am1757 & Am0707)) were incubated with three different

substrates (representing the 3 indicated types of sialic acid bond) for 30 minutes, after which the amount of free sialic acid was measured.

Figure 3: Optimal conditions for Am0707 and Am1757 were determined. Conditions affecting the enzymatic activity of the sialidases, including ions, NaCl, and pH were investigated using 2-3 sialyllactose substrate, with free sialic acids being quantitated after 15 minutes incubation.

Figure 4: Combined sialidase treatment of glycoproteins is more efficient than using single sialidases. Sialidases were incubated with 0.5 µg fetuin for 60 minutes after which they were separated using SDS-PAGE. Smix contains both Am0707 and Am1757, while a *C. perfringens* sialidase was used as a benchmark comparison. All reactions took place at 37°C, in 20 mM Tris-HCl pH 6.8, except for the benchmark product which was incubated according to manufacturer's instructions.

Figure 5: GVS_Smix is superior to current commercially available sialidases. Several existing commercial sialidases from New England Biolabs (first 3 columns in each set) were tested alongside the AM1757+Am0707 mixture (GVS_Smix) and each of the Am1757 and Am0707 enzymes individually. Each was incubated with specific sialidase substrates (representing the 3 indicated types of sialic acid bond) for 30 min at 37°C before being analyzed. FU (Fluorescence units) represent the amount of sialic acids liberated.

Figure 6: GVS_Smix can fully asialylate fetuin. SDS-PAGE and SNA blotting show that Smix (the AM1757+Am0707 mixture) as well as two New England Biolabs products (NEB A and NEB O) can fully asialylate the 2-3 and 2-6 sialic acid bonds in fetuin. The two 2-3 specific enzymes Am1757 and New England Biolabs NEB S were not able to fully asialylate fetuin.

Figure 7: The GVS_Smix (the AM1757+Am0707 mixture; first set of bars) releases sialic acids from native proteins at a level similar to, or better, than all three tested commercial products (the last three sets of bars). The Am1757 enzyme alone is also shown (second set of bars). Proteins were mixed with sialidases in their respective buffer and incubated for 15 minutes before addition of sialic acid development buffer. All incubations took place at 37°C in the respective buffers (for GVS-Smix this was 20 mM Tris-HCl pH 6.8).

Figure 8: Recombinant expression of O-glycosidases. (A) O-glycosidase from *S. oralis* and *B. bifidum* were expressed, affinity purified, and separated on SDS-PAGE. (B) Stability assay of *S. oralis* O-glycosidase stored at 4°C for 0-6 days.

Figure 9: O-glycosidases can hydrolyze pNP-labeled O-glycans of types Core 1, 2 and 3. Different concentrations of O-glycosidases from *S. oralis* (*S.o*) and *B. bifidum* (*B.b*) were incubated with (A) Core 1, (B) Core 2, and (C) Core 3 O-glycans, and release of pNP was measured as change in absorbance AT 405 nm.

Figure 10: The *S. oralis* glycosidase has higher activity in basic pH supplemented with MgCl₂. The O-glycosidases were incubated at different pH (A), with different ions (B), and different concentrations of MgCl₂ (C) to determine the optimal buffer.

Figure 11: O-glycosidase activity is time and dose-dependent. TNFαR (1μg) was incubated with the GVS_Smix sialidase mixture only (lanes labelled TNFaR (Smix)) as a control, or with the GVS_Smix sialidase mixture in combination with the different amounts of O-glycosidase from *S. oralis* (So) that are shown (“+5 ug So etc”). The denoted amounts of added O-glycosidases rely on values from the nanodrop. However, due to fragmentation, the actual amount full-length protein added is closer to ca 10-20% of the written value. These samples were incubated in 20 mM Tris-HCl pH 8.0, supplemented with 2 mM CaCl₂, but no MgCl₂.

Figure 12: O-glycosidases can act upon all investigated native glycoproteins. Different native proteins (P) were incubated with the GVS_Smix mixture of sialidases (Smix) and O-glycosidases from *S. oralis* (*So-glyk*) and *B. bifidum* (*Bb-glyk*). Incubations were o/n, after which products were analyzed on SDS-PAGE.

Figure 13: The *S. oralis* O-glycosidase compares favorably with commercially available products for action against native glycoproteins. Native TNFαR was incubated with GVS_Smix sialidase mixture only (lanes labelled TNFaR (Smix)) as a control, or with the GVS_Smix sialidase mixture in combination with the different amounts of O-glycosidase from *S. oralis* (So) that are shown (“+ So (0.5ug)” etc) or with a commercially available O-glycosidase (NEB) at the quantity shown (“+NEB (3ul)” etc) in combination with a sialidase supplied with it by the manufacturer. All enzymes were incubated with TNFαR in their respective buffers for either 1 hour or 16 hours. The highest dose of the *S. oralis* glycosidase (“So (0.5ug)”) roughly equals 0.3 ul of the commercial O-glycosidase

(NEB) in molarity. The distinct band of low molecular weight (top of gel) in the NEB treated samples is the accompanying sialidase.

Figure 14: Glycan composition influences activity. Incubation of the substrate with sialidase (S) and/or galactosidase (G) before addition of O-glycosidase (So); incubating o/n
5 at 37°C in optimal buffers (20 mM Tris-HCl pH 8.0).

Figure 15: Finalized product is competitive. Using a enzyme:substrate 1:40 ratio of the O-glycosidase, as well as enzyme substrate ratio of 1:40+1:40 for the Am1757+Am0707 in the sialidase mixture, all three enzymes in 20 mM Tris-HCl pH 6.8 (lanes labeled Smix / O-glyk), the activity of the *S. oralis* O-glycosidase was compared to
10 that of the commercial O-glycosidase + sialidase combination (lanes labelled NEB (O-glyk kit). O-glycan removal from TNFaR by the *S. oralis* O-glycosidase + sialidase mix (“+GVS”) was verified with lectin blotting, using ConA (N-linked glycans) and jacalin (O-linked glycans), untreated TNFaR as a control.

Figure 16: The O-glycosidase + sialidase bundle (Smix / O-glyk) is active against
15 a wide variety of glycoproteins. The enzymes added in their final concentration and formulation (1:40) in 20 mM Tris-HCl pH 6.8 were incubated with the glycoproteins for 4 h at 37°C and then separated on SDS-PAGE.

Figure 17: Specific sialic acid bonds influence O-glycosidase activity. The enzymes (Am1757, Am0707 or Smix) added in their final concentration and formulation
20 (1:40) in 20 mM Tris-HCl pH 6.8 were incubated with the TNFaR for 15 min – 4 h at 37°C and then separated on SDS-PAGE.

Figure 18: The *S. oralis* O-glycosidase excels at least in part due to potent sialidases. *S. oralis* O-glycosidase (GVS O-glyk) or commercial O-glycosidase (NEB O-glyk) were incubated with the GVS_Smix sialidase mixture (GVS sialidase) or with
25 commercial sialidases (NEB A, NEB S, NEB O). Appropriate buffers were used (e.g. 20 mM Tris-HCl pH 6.8 for GVS), and each set of enzymes was incubated with Enbrel (Etanercept) for 4 h at 37°C, before separation of products on SDS-PAGE.

Figure 19: 2-3 bonded sialic acids limit the efficiency of an O-glycoprotein specific endoprotease (LS). Concurrent incubation of LS with a set of diverse sialidases
30 for 30 min - 20 h, using Enbrel as a glycoprotein substrate, revealed the higher efficiency in the presence of the 2-3 specific sialdiase 1757, or with the Mix (0707 + 1757), while the

broad spectrum sialidase 0707 was not necessary for seemingly full activity of LS, thus suggesting that 2-6 (and 2-8) bonds are not a concern for LS activity.

Brief Description of the Sequences

5 SEQ ID NOs: 1, 2 and 3 are each an amino acid sequence of a sialidase isolated from *Akkermansia muciniphila*. SEQ ID NO: 1 is the wildtype sequence including a signal motif at the N terminus. SEQ ID NO: 2 is the wildtype sequence with signal motif removed. SEQ ID NO: 3 is identical to SEQ ID NO: 2, except it includes an additional N terminal Methionine and a C-terminal linker + His₆ tag. Any sequence comprising the
10 sequence of SEQ ID NO: 2 (including each of SEQ ID NOs: 1 to 3) may be referred to herein as Am0707.

 SEQ ID NOs: 4, 5 and 6 are each an amino acid sequence of another sialidase isolated from *Akkermansia muciniphila*. SEQ ID NO: 4 is the wildtype sequence including a signal motif at the N terminus. SEQ ID NO: 5 is the wildtype sequence with signal motif
15 removed. SEQ ID NO: 6 is identical to SEQ ID NO: 5, except it includes an additional N terminal Methionine and a C-terminal linker + His₆ tag. Any sequence comprising the sequence of SEQ ID NO: 5 (including each of SEQ ID NOs: 4 to 6) may be referred to herein as Am1757.

 SEQ ID NOs: 7, 8, 9 and 10 are each an amino sequence of an O-glycosidase
20 isolated from *S. oralis*. SEQ ID NO: 7 is the wildtype sequence including a signal motif at the N terminus and an LPXTG wall anchor motif at the C terminus. SEQ ID NO: 8 is the wildtype sequence with signal motif removed. SEQ ID NO: 9 is the wildtype sequence with signal motif and the wall anchor motif removed. SEQ ID NO: 10 is identical to SEQ ID NO: 9, except it includes an additional N terminal Methionine and a C-terminal linker +
25 His₆ tag. Any sequence comprising the sequence of SEQ ID NO: 9 (including each of SEQ ID NOs: 7 to 10) may be referred to herein as “O-glyk” or “So”.

 SEQ ID NO: 11 is an amino acid sequence of an O-glycoprotein-specific endoprotease.

 SEQ ID NO: 12 is the amino acid sequence of an exemplary polypeptide having O-
30 glycoprotein-specific endoprotease activity. Relative to SEQ ID NO: 11 it includes an additional N terminal Methionine and a C-terminal linker + His₆ tag. The polypeptide consisting of this sequence may be referred to herein as LS.

SEQ ID NO: 13 is an exemplary nucleic acid sequence encoding the polypeptide consisting of the amino acid sequence of SEQ ID NO: 10.

Detailed Description of the Invention

5 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, and is not intended to be limiting. All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in
10 their entirety.

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.

This specification is particularly concerned with polypeptides which are sialidases,
15 O-glycosidases and O-glycoprotein-specific endoproteases. General uses of the term polypeptide may thus be applied to each of these types of enzyme.

General polypeptide features

A “polypeptide” is used herein in its broadest sense to refer to a compound of two
20 or more subunit amino acids, amino acid analogs, or other peptidomimetics. The term “polypeptide” thus includes short peptide sequences and also 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 amino acids, including both D or L optical isomers, and amino acid analogs and
25 peptidomimetics.

A polypeptide may be produced by suitable method, including recombinant or synthetic 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
30 produced by transforming a cell, typically a bacterial cell, with a nucleic acid molecule or vector which encodes said polypeptide. Production of polypeptides by expression in bacterial host cells is described below and is exemplified in the Examples. The invention

provides nucleic acid molecules and vectors which encode a polypeptide of the invention. The invention also provides a host cell comprising such a nucleic acid or vector. An exemplary polynucleotide molecules encoding a polypeptide disclosed herein is provided as SEQ ID NO: 13. This sequence includes at the 3' end a codon for the N terminal
5 methionine (ATG) and, prior to the stop codon (TAA) at the 5' end, codons for a GSGLE 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.

The terms "nucleic acid molecule" and "polynucleotide" are used interchangeably herein and refer to a polymeric form of nucleotides of any length, either
10 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. A polynucleotide of the invention encodes a polypeptide of the invention and may be provided in isolated or substantially
15 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. A nucleic acid sequence which "encodes" a selected polypeptide is a nucleic acid molecule which is transcribed (in the case of DNA) and
20 translated (in the case of mRNA) into a polypeptide in vivo 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
25 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.

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
30 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 operably linked to the inserted sequence, thus allowing for expression of the polypeptide of

the invention in vivo. 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 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.

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 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.

The invention also includes cells that have been modified to express a polypeptide of the invention. Such 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.

A polypeptide may be derivatised or modified to assist with their production, isolation or 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 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 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, preferably at the C-terminus. It may be joined directly to the polypeptide or joined indirectly by any suitable linker sequence, such as 3, 4 or 5 amino acids. The linker may typically consist predominantly of glycine and serine residues. The histidine tag
5 typically consists of six histidine residues, although it can be longer than this, typically up to 7, 8, 9, 10 or 20 amino acids or shorter, for example 5, 4, 3, 2 or 1 amino acids.

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
10 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
15 purified. Where a polypeptide is provided in a composition 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.

20 A polypeptide (or mixture thereof) may be provided in lyophilised form, suitable for reconstitution in aqueous solution prior to use. The lyophilised composition has improved stability enabling 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
25 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.

30 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, 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
5 example, the polypeptide may be immobilised on sepharose.

Sialidases

Functional features

Besides using chemical and genetic approaches to modify glycan chains, several
10 enzymes (glycosidases) can act upon the bonds linking sialic acids to other glycans. These enzymes, termed sialidases or neuraminidases, show a high degree of specificity for particular types of sialic acid bond. Three distinct bond types are commonly found within human glycoproteins, with alpha(2-3) bonds being the dominant form, followed by alpha(2-6) and alpha(2-8). These bond types may be referred to herein as 2-3, 2-6 and 2-8
15 bonds for simplicity. A 2-3 bond means that the carbon atom at position number 2 of the sialic acid hexose is joined, via an oxygen atom, to the carbon at position 3 of the hexose of the linked glycan. Correspondingly, a 2-6 bond or a 2-8 bond means that the join is to position 6 or position 8 of the hexose of the linked glycan, respectively.

Most known sialidases are either specific for the 2-3 bond (cleaving it with very
20 high activity), or are able to cleave a wider range of bonds, typically all of 2-3, 2-6 and 2-8 bonds. These different types of sialidase may be referred to as narrow spectrum or broad spectrum, respectively. Broad spectrum sialidases typically exhibit high activity against 2-3 bonds, with decreasing activity against 2-6, and very low activity against 2-8 bonds. Enzymes which cleave 2-8 bonds efficiently are comparatively rare (even unknown) in the
25 field.

The enzymatic activity of a sialidase may be assessed by any suitable method, such as those described in the Examples. A suitable method may include incubating a known or suspected sialidase with a standard sialidase substrate, such as one or more small molecules which collectively comprise 2-3, 2-6 and 2-8 type bonds. Such small molecules
30 include 2-3'-sialyllactose, 2-6'-sialyllactose, and colominic acid (2-8'). Sialidase activity on such molecules will result in free sialic acids, which may be quantified by routine methods. Alternatively sialidase activity may be assessed using a glycoprotein as

substrate. Any resulting cleavage products may be detected and quantified using routine methods such as SDS-PAGE or lectin blots.

Structural features

5 The present inventors have identified and characterized several sialidases from the commensal gut bacterium *Akkermansia muciniphila*. One of the sialidases, referred to herein as Am0707, has unexpectedly high activity against 2-8 bonds, but can also cleave 2-3 and 2-6 bonds. It may thus be considered a broad spectrum sialidase. Another sialidase, referred to herein as Am1757, has exclusively high activity against 2-3 bonds. It may thus
10 be considered a narrow spectrum sialidase.

 The full wildtype primary structure (amino acid sequence) of the first sialidase (Am0707) is shown in SEQ ID NO: 1. The sequence with signal motif removed is shown in SEQ ID NO: 2. The first sialidase may comprise, consist essentially, or consist of the sequence of SEQ ID NO: 2, and is typically no longer than 400 amino acids.

15 The full wildtype primary structure (amino acid sequence) of the second sialidase (Am1757) is shown in SEQ ID NO: 4. The sequence with signal motif removed is shown in SEQ ID NO: 5. The second sialidase may comprise, consist essentially, or consist of the sequence of SEQ ID NO: 5, and is typically no longer than 600 amino acids..

 Alternatively, the said first and/or said second sialidase may each independently be
20 replaced by a variant of each thereof, provided that enzymatic activity is retained. A variant of a said sialidase may comprise, consist essentially, or consist of a variant of the amino acid sequence of sequence of SEQ ID NO: 2 or 5, respectively, which is at least 50% identical to said amino acid sequence. The variant sequence may be at least 60%, at least 70%, at least 80%, at least, 85%, at least 90%, at least 95%, at least 98% or at least
25 99% identical to said amino acid sequence. The identity level is preferably at least 85% or higher. Identity relative to a sequence can be measured over a region of at least 100, at least 200, at least 300, at least 350, at least 400, or at least 500 or more contiguous amino acids of the sequence, or more preferably over the full length of the sequence. A variant is typically of a length which is no more than 50 amino acids longer or shorter than the
30 reference sequence, and is preferably of approximately (or exactly) the same length as the reference sequence.

Amino acid identity may be calculated using any suitable algorithm. For example the PILEUP and BLAST algorithms can be used to calculate identity or line up sequences (such as identifying equivalent or corresponding sequences (typically on their default settings), for example as described in Altschul S. F. (1993) *J Mol Evol* 36:290-300;

5 Altschul, S, F *et al* (1990) *J Mol Biol* 215:403-10. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pair (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of

10 the same length in a database sequence. T is referred to as the neighbourhood word score threshold (Altschul *et al*, supra). These initial neighbourhood word hits act as seeds for initiating searches to find HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extensions for the word hits in each direction are halted when: the cumulative

15 alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring

20 matrix (see Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

The BLAST algorithm performs a statistical analysis of the similarity between two sequences; see e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum

25 probability (P(N)), which provides an indication of the probability by which a match between two polynucleotide or amino acid sequences would occur by chance. For example, a sequence is considered similar to another sequence if the smallest sum probability in comparison of the first sequence to the second sequence is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably

30 less than about 0.001. Alternatively, the UWGCG Package provides the BESTFIT program which can be used to calculate identity (for example used on its default settings) (Devereux *et al* (1984) *Nucleic Acids Research* 12, 387-395).

The sequence of a sialidase may comprise a variant of the respective SEQ ID NO in which modifications, such as amino acid additions, deletions or substitutions are made relative to the sequence said SEQ ID NO. 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, 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 A1 below. Where amino acids have similar polarity, this can be determined by reference to the hydropathy scale for amino acid side chains in Table A2. A sequence of a sialidase of the invention may comprise a variant of the respective SEQ ID NO in which upto 10, 20, 30, 40, 50 or 60 conservative substitutions are made.

Table A1 - Chemical properties of amino acids

Ala (A)	aliphatic, hydrophobic, neutral	Met (M)	hydrophobic, neutral
Cys (C)	polar, hydrophobic, neutral	Asn (N)	polar, hydrophilic, neutral
Asp (D)	polar, hydrophilic, charged (-)	Pro (P)	hydrophobic, neutral
Glu (E)	polar, hydrophilic, charged (-)	Gln (Q)	polar, hydrophilic, neutral
Phe (F)	aromatic, hydrophobic, neutral	Arg (R)	polar, hydrophilic, charged (+)
Gly (G)	aliphatic, neutral	Ser (S)	polar, hydrophilic, neutral
His (H)	aromatic, polar, hydrophilic, charged (+)	Thr (T)	polar, hydrophilic, neutral
Ile (I)	aliphatic, hydrophobic, neutral	Val (V)	aliphatic, hydrophobic, neutral
Lys (K)	polar, hydrophilic, charged(+)	Trp (W)	aromatic, hydrophobic, neutral
Leu (L)	aliphatic, hydrophobic, neutral	Tyr (Y)	aromatic, polar, hydrophobic

Table A2 - Hydropathy scale

	<u>Side Chain</u>	<u>Hydropathy</u>
20	Ile	4.5
	Val	4.2
	Leu	3.8
	Phe	2.8
25	Cys	2.5
	Met	1.9
	Ala	1.8
	Gly	-0.4
	Thr	-0.7
	Ser	-0.8
30	Trp	-0.9
	Tyr	-1.3
	Pro	-1.6

	His	-3.2
	Glu	-3.5
	Gln	-3.5
5	Asp	-3.5
	Asn	-3.5
	Lys	-3.9
	Arg	-4.5

Alternatively, a sialidase may be replaced by a shorter fragment of the respective SEQ ID NO or of a variant thereof as described above. The fragments may be described as
 10 a truncated form of said SEQ ID NO which retains enzymatic activity. Such fragments are shorter than the corresponding SEQ ID NO and are typically at least 100, 150, 200, 250, 300, 350, 400, 450 or 500 amino acids in length.

Any sialidase described herein may optionally include an additional methionine at the N terminus and/or a histidine or other tag at the C terminus. Such additional sequences
 15 may aid with expression and/or purification. A histidine tag preferably consists of six histidine residues. The histidine tag is preferably linked to the C terminus by a linker, which is typically a short sequence of amino acids, such as 3 – 5 amino acids. The linker typically consists predominantly of glycine and serine residues, and may preferably include the sequence GSG. For example GSG and GSGLE are suitable linkers.

20 In summary therefore, a first sialidase is:

- (a) a polypeptide comprising or consisting of an amino acid sequence of SEQ ID NO: 2;
- (b) a polypeptide comprising or consisting of an amino acid sequence which is at least 85% identical to the amino acid sequence of SEQ ID NO: 2 or
- 25 (c) a polypeptide comprising or consisting of an amino acid sequence which is a fragment of the sequence of SEQ ID NO: 2 or a fragment of an amino acid which is 85% identical to the amino acid sequence of SEQ ID NO: 2;

and a second sialidase is:

- (d) a polypeptide comprising or consisting of an amino acid sequence of
 30 SEQ ID NO: 5;
- (e) a polypeptide comprising or consisting of an amino acid sequence which is at least 85% identical to the amino acid sequence of SEQ ID NO: 5 or
- (f) a polypeptide comprising or consisting of an amino acid sequence which is a fragment of the sequence of SEQ ID NO: 5 or a fragment of an amino
 35 acid which is 85% identical to the amino acid sequence of SEQ ID NO: 5;

optionally wherein said first and/or said second sialidase includes an additional methionine at the N terminus and/or a histidine tag at the C terminus, which tag may be joined to the C terminus by a linker.

An exemplary first sialidase is the polypeptide consisting of the amino acid sequence of SEQ ID NO: 3. An exemplary second sialidase is the polypeptide consisting of the amino acid sequence of SEQ ID NO: 6.

Sialidase compositions

A sialidase composition comprises at least one sialidase, preferably in substantially isolated or purified form. As in the general disclosure relating to polypeptides set out above, this typically means isolated from the majority of the other components present in a cellular extract from a cell in which the sialidase was expressed. By substantially purified, it will be understood that the sialidase 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 sialidase may be mixed with carriers, diluents or preservatives which will not interfere with the intended purpose of the sialidase and still be regarded as substantially isolated or purified. The sialidase composition may comprise an additional active component, such as another sialidase or another enzyme, in which case each said component will individually be purified to a high level of homogeneity prior to mixing in an appropriate ratio for the intended purpose of each. In a preferred sialidase composition of the invention, the composition comprises a first sialidase and a second sialidase which are each purified to at least 90% homogeneity and are present at a 1:1 ratio to each other. Such a composition may include an additional active component, such as another enzyme which is not a sialidase. The other enzyme may be a protease and/or a glycosidase. The protease is preferably an O-glycoprotein-specific endoprotease. The glycosidase is preferably an O-glycosidase. Both types of enzyme are discussed in more detail below.

Where a sialidase composition includes an active component which is not a sialidase, the preferred ratio of total sialidase content (e.g. first plus second sialidase) relative to the other enzyme will be 1:1. For example, if a composition includes 2000 units of another enzyme, it will also include 2000 units of sialidase, in which if there are

two sialidases, then said 2000 units comprises 1000 units of the first sialidase and 1000 units of the second sialidase.

A sialidase composition (as with a polypeptide in general) may be provided in lyophilised form, suitable for reconstitution in aqueous solution prior to use. The lyophilised composition has improved stability enabling longer storage of the sialidase(s). A method of preparing a sialidase composition in lyophilised form, comprising freeze-drying one or more sialidases in a suitable buffer, such as Tris-buffered saline (TBS), is provided herein. The buffer preferably comprises a low concentration of NaCl, typically upto 300mM, 250mM, 200mM, or 150mM. The NaCl concentration is preferably around 150mM, such as between 125mM and 175mM. A sialidase is typically substantially purified prior to freeze-drying. The resulting lyophilised form of the composition is also provided. A method of preparing a sialidase composition which is a solution, comprising providing the composition in lyophilised form and reconstituting with a suitable carrier or diluent, such as water, is also provided.

The present inventors determined that a first sialidase has unusually high activity against 2-8 bonds. Accordingly the present invention provides a composition which comprises a first sialidase which is independently selected from:

- (a) a polypeptide comprising or consisting of an amino acid sequence of SEQ ID NO: 2;
 - (b) a polypeptide comprising or consisting of an amino acid sequence which is at least 85% identical to the amino acid sequence of SEQ ID NO: 2 or
 - (c) a polypeptide comprising or consisting of an amino acid sequence which is a fragment of the sequence of SEQ ID NO: 2 or a fragment of an amino acid which is 85% identical to the amino acid sequence of SEQ ID NO: 2;
- optionally wherein said first sialidase includes an additional methionine at the N terminus and/or a histidine tag at the C terminus, which tag may be joined to the C terminus by a linker.

Said composition may be for use in a method of cleaving 2-8 sialic acid bonds, preferably with high efficiency. An example of such a composition comprises the sialidase consisting of the amino acid sequence of SEQ ID NO: 3.

The present inventors also determined that a combination of a first sialidase (Am0707) and a second sialidase (Am1757) hydrolyses 2-3, 2-6, and 2-8 bonds with

unusually high efficiency, thereby permitting the efficient removal of substantially all (typically >90%) of the sialic acids of any glycoprotein. The combination was also surprisingly effective against glycoproteins in a native (that is non-denatured) state. Accordingly the present invention provides a composition as described above comprising a first sialidase, which composition further comprises a second sialidase which is independently selected from:

- (d) a polypeptide comprising or consisting of an amino acid sequence of SEQ ID NO: 5;
- (e) a polypeptide comprising or consisting of an amino acid sequence which is at least 85% identical to the amino acid sequence of SEQ ID NO: 5 or
- (f) a polypeptide comprising or consisting of an amino acid sequence which is a fragment of the sequence of SEQ ID NO: 5 or a fragment of an amino acid which is 85% identical to the amino acid sequence of SEQ ID NO: 5.

optionally wherein said second sialidase includes an additional methionine at the N terminus and/or a histidine tag at the C terminus, which tag may be joined to the C terminus by a linker. The first and second sialidase may preferably be present in a 1:1 ratio relative to each other.

Said composition may be for use in a method of completely asialylating a glycoprotein, or of cleaving >90% of the sialic bonds in a glycoprotein, preferably with high efficiency. The glycoprotein is preferably in a native state. That is, it has not been subjected to any form of denaturing conditions.

An example of a sialidase composition of the invention comprises the sialidase consisting of the amino acid sequence of SEQ ID NO: 3 and the sialidase consisting of the amino acid sequence of SEQ ID NO: 6, preferably in a 1:1 ratio.

The sialidase activity of a sialidase composition may be assessed using the same methods as described above for individual sialidases. However, it is preferably assessed using a non-denatured glycoprotein as substrate. The results may be compared to those obtained in the same assay when the substrate is contacted with an exemplary sialidase or mixture thereof, such as a 1:1 mixture of the polypeptide consisting of the amino acid sequence of SEQ ID NO: 3 and the polypeptide consisting of the amino acid sequence of SEQ ID NO: 6. A unit of such a sialidase mixture is typically the amount required to hydrolyse sialic acids from $\geq 90\%$ of 1 μg glycoprotein (fetuin) when incubated in 20 mM

Tris pH 6.8 at 37 °C for 2 h at 37°C as monitored by SDS-PAGE. This is considered to represent high efficiency.

O-glycosidase

5 The present inventors have also identified and characterised an O-glycosidase from the commensal bacterium *Streptococcus oralis*, residing in the oral tract of mammals, which efficiently hydrolyses O-linked glycans particularly when used in combination with a sialidase composition as described above. The O-glycosidase may be referred to herein as “O-glyk” or “So”. The wild-type sequence of O-glyk is provided as SEQ ID NO: 7,
10 which includes a signal sequence and an LPTXG cell wall anchor motif. The wild-type sequence of O-glyk lacking the signal sequence is provided as SEQ ID NO: 8. The wild-type sequence of O-glyk lacking the signal sequence and the C terminal part of the cell wall anchor motif is provided as SEQ ID NO: 9. These sequences can be optionally modified to include an additional methionine at the N terminus and/or a histidine or other
15 tag at the C terminus. Such additional sequences may aid with expression (e.g. in *E. coli*) and/or purification. A histidine tag preferably consists of six histidine residues. The histidine tag is preferably linked to the C terminus by a linker, which is typically a short sequence of amino acids, such as 3 – 5 amino acids. The linker typically consists
20 predominantly of glycine and serine residues, and may preferably include the sequence GSG. For example GSG and GSGLE are suitable linkers. An exemplary O-glyk sequence having an additional methionine at the N terminus and a GSGLE linker and His6 tag at the C terminus is provided as SEQ ID NO: 10. Any reference to “O-glyk” or “So” in the present disclosure may mean any of SEQ ID NOs: 7, 8, 9 or 10, but preferably refers to a polypeptide which comprises or consists of the amino acid sequence of SEQ ID NO: 9, and
25 is typically no longer than 2070 amino acids. Most preferred is a polypeptide which consists of the amino acid sequence of SEQ ID NO: 10.

 The present inventors also discovered that the action of a sialidase composition as described above also enhances the activity of other O-glycosidases. Thus,
the present invention also provides a method of modifying a glycoprotein comprising
30 contacting a sample of glycoprotein both with a sialidase composition as described above and with an O-glycosidase. The present invention also provides a sialidase composition as described above which optionally also comprises an O-glycosidase. In said method and

said composition, said O-glycosidase may comprise or consist of the amino acid sequence of SEQ ID NO: 9 or may be any other O-glycosidase, such as an enzyme obtained from a gut bacterium e.g. *Enterococcus faecalis*. A preferred O-glycosidase from *E. faecalis* is a polypeptide having the amino acid sequence of Uniprot entry B5UB72 version 22. Other
5 suitable O-glycosidases include those described in WO2009129086, particularly EngEF, EngPA and truncated EngAA as described on page 7 and shown in Figure 5 of WO2009129086.

In any of the disclosures herein, the O-glycosidase comprising the sequence of SEQ ID NO: 9 may be replaced by a variant thereof, provided that enzymatic activity is
10 retained. A variant of the O-glycosidase may comprise, consist essentially, or consist of a variant of the amino acid sequence of sequence of SEQ ID NO: 9. The variant of said SEQ ID NO may be defined as set out above with respect to the sialidases, except that the relevant enzymatic activity to be retained is hydrolytic activity against O-glycans.

Alternatively, the O-glycosidase may be replaced by a shorter fragment of SEQ ID
15 NO: 9 or of a variant thereof as described above. The fragments may be described as a truncated form of said SEQ ID NO which retains enzymatic activity. Such fragments are shorter than SEQ ID NO: 3 and are typically at least 300, 400, 500, 600, 800, 1000, 1200, 1300, 1400 or 1500 amino acids in length.

Any O-glycosidase described herein may optionally include an additional
20 methionine at the N terminus and/or a histidine or other tag at the C terminus. Such additional sequences may aid with expression and/or purification. A histidine tag preferably consists of six histidine residues. The histidine tag is preferably linked to the C terminus by a linker, which is typically a short sequence of amino acids, such as 3 – 5 amino acids. The linker typically consists predominantly of glycine and serine residues,
25 and may preferably include the sequence GSG. For example GSG and GSGLE are suitable linkers. An exemplary O-glycosidase of this type consists of the amino acid sequence of SEQ ID NO: 10.

The enzymatic activity of an O-glycosidase may be assessed by any suitable method, such as those described in the Examples. A suitable method may include
30 incubating a known or suspected O-glycosidase with a standard substrate, such as one or more small molecules which collectively comprise O-glycan core regions. Such small molecules include 4-Methylumbelliferone (4MU) substrates and pNP-substrates, with the

release of pNP indicating activity. Alternatively activity may be assessed using a glycoprotein as substrate. Any resulting cleavage products may be detected and quantified using routine methods such as SDS-PAGE or lectin blots. Where a glycoprotein is used as the substrate pre-treatment (or simultaneous treatment) with a sialidase composition as described above may be required. The results may be compared to those obtained in the same assay when the substrate is contacted with an exemplary O-glycosidase, such as a polypeptide consisting of the acid sequence of SEQ ID NO: 10. One unit of the polypeptide of SEQ ID NO: 10 is defined as the amount required to remove O-glycans from > 90% of 1 µg of TNFaR in combination with one unit of a sialidase mixture in 20 mM Tris buffer pH 6.8, in 2 hours at 37°C as monitored by SDS-PAGE (preferred sialidase mixtures are as described above). A test polypeptide preferably achieves a similar level of activity when present in the same amount. Exemplary assays are also described in the Examples.

A composition comprising an O-glycosidase as described above may be provided in solution or in lyophilised form for reconstitution in solution. The O-glycosidase may be lyophilized in Tris buffer saline pH 7.6.

O-glycan specific endoprotease

The present inventors also discovered that the action of a sialidase composition as described above enhances the activity of O-glycan specific endoproteases, in particular an O-glycan specific endoprotease comprising the amino acid sequence of SEQ ID NO: 11 which is typically no longer than 375 amino acids, and which is preferably a polypeptide consisting of the amino acid sequence of SEQ ID NO: 12.

Thus, the present invention also provides a method of modifying a glycoprotein comprising contacting a sample of glycoprotein both with a sialidase composition as described above and with an O-glycan specific endoprotease. The present invention also provides a sialidase composition as described above which optionally also comprises an O-glycan specific endoprotease. In said method and said composition, said O-glycan specific endoprotease may be that of SEQ ID NO: 12.

In any of the disclosures herein, the O-glycan specific endoproteases of SEQ ID NO: 11 may be replaced by a variant thereof, provided that enzymatic activity is retained. A variant of the endoprotease may comprise, consist essentially, or consist of a variant of

the amino acid sequence of sequence of SEQ ID NO: 11. The variant of said SEQ ID NO may be defined as set out above with respect to the sialidases, except that the relevant enzymatic activity to be retained is hydrolytic activity against O-glycoproteins.

Alternatively, the endoprotease may be replaced by a shorter fragment of SEQ ID NO: 3 or of a variant thereof as described above. The fragments may be described as a truncated form of said SEQ ID NO which retains enzymatic activity. Such fragments are shorter than SEQ ID NO: 11 and are typically at least 100, 150, 200, 250, 300, 350, 400, 450 or 500 amino acids in length.

Any endoprotease described herein may optionally include an additional methionine at the N terminus and/or a histidine or other tag at the C terminus. Such additional sequences may aid with expression and/or purification. A histidine tag preferably consists of six histidine residues. The histidine tag is preferably linked to the C terminus by a linker, which is typically a short sequence of amino acids, such as 3 – 5 amino acids. The linker typically consists predominantly of glycine and serine residues, and may preferably include the sequence GSG. For example GSG and GSGLE are suitable linkers. An exemplary endoprotease of this type consists of the amino acid sequence of SEQ ID NO: 12.

The enzymatic activity of an endoprotease may be assessed by any suitable assay. For example, a standard O-glycoprotein substrate, such as an IgA molecule, may be incubated with a test polypeptide. The starting materials and the reaction products may then be analysed by SDS-PAGE and/or mass spectrometry to determine the presence of cleavage products (if any) and if required also to further characterise those products. A glycoprotein substrate which is not O-glycosylated, such as an IgG1 molecule, may be used as a negative control. The results may be compared to those obtained in the same assay when the substrate is contacted with an exemplary polypeptide, such as a polypeptide consisting of the amino acid sequence of SEQ ID NO: 12.

A composition comprising an O-glycan specific endoprotease as described above may be provided in solution or in lyophilised form for reconstitution in solution. The O-glycan specific endoprotease may be lyophilized in Tris buffer saline pH 7.6.

30

Methods of use

The present invention provides any methods in which a sample is incubated with a composition of the invention under conditions suitable for the enzymes in said composition to act upon any substrates that are present. Said methods may optionally include an analysis of the resulting products. Said analysis may include the separation and/or
5 detection and/or isolation of the products by any suitable means, including SDS-PAGE, HPLC, lectin blotting, ELISA or mass spectrometry.

Suitable conditions include incubation with a composition of the invention for at least 20 minutes, 30 minutes, 40 minutes, 50 minutes, 60 minutes, 70 minutes, 80 minutes, 90 minutes or 120 minutes, 3 hours, 5 hours, 10 hours, 12 hours, or overnight. Incubation
10 preferably takes place at room temperature, more preferably at approximately 20°C, 25°C, 30°C, 35°C, 40°C or 45°C, and most preferably at approximately 37°C. The methods may be carried out under any suitable pH. Suitable pH values include, for example, a pH of around 3.0, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9 or 9.5. Preferred pH is in the range 5.6 to 6.8. The methods may be conducted in any suitable buffer, such as tris buffered
15 saline (TBS) or phosphate buffered saline (PBS). The buffer preferably comprises a low concentration of NaCl, typically no more than 300mM, 250mM, 200mM, or 150mM. The NaCl concentration is preferably around 150mM, such as between 125 and 175mM. The approximate ratio of the enzymes in the composition of the invention to the protein content of the sample may be 1:1, 2:1, 4:1, 6:1, 10:1, 15:1, 20:1, 1:2, 1:4, or 1:6, 1:10,
20 1:15, 1:20, 1:40, 1:50 or 1:100.

The following are particularly preferred methods of the invention:

A method for the modification of a glycoprotein, the method comprising contacting a sample containing the glycoprotein with a composition of the invention and optionally analyzing the resulting products. Said analysis may comprise separating and/or detecting
25 and/or isolating the products by any suitable method, including SDS-PAGE, HPLC, lectin blotting, ELISA or mass spectrometry.

In a particular embodiment, the method may comprise contacting a sample with a composition of the invention which comprises only sialidases, optionally separating the products, and then contacting the said products with another enzyme, such as a protease
30 and/or a glycosidase. This method may be described as a “pre-treatment” of the sample to remove sialic acid before the other enzyme is added. In a variation of this embodiment, the other enzyme may be added to the sample separately but simultaneously with the sialidase

composition, and may be described as a “co-treatment”. In another variation, the other enzyme is present in the sialidase composition. The other enzyme is preferably an O-glycan specific endoprotease or an O-glycosidase, for example as described herein.

The following Examples illustrate the invention:

5

Example 1 - Sialidases

Materials & Methods

Expression and purification of sialidases

10 Genes (Am0705, Am0707, Am1757, Am2058) identified in *Akkermansia muciniphila* were codon-optimized to express well in *E. coli* in the vector pET21a(+). The vector was transformed into BL21(DE3) Star cells. *E. coli* was routinely cultured in LB at 37°C, 200 rpm. In the presence of the plasmid, 100 µg/mL ampicillin was added. After o/n incubation, cultures were diluted 1:20 in fresh LB(amp), and grown until OD₆₂₀ ~ 0.7-0.8, 15 after which recombinant protein expression was induced by addition of 1 mM IPTG, and the expression continued for 5 hours before the cells were collected and frozen. Frozen cells were thawed and resolved in His binding buffer (20 mM NaP pH 7.4, 500 mM NaCl, 20 mM imidazole), and sonicated for release of intracellular proteins. Cell debris was removed by centrifugation. Sterile filtered supernatant was affinity purified on a nickel 20 column, and re-buffered to 20 mM Tris-HCl pH 8.0 on a PD-25 column. Concentration of the proteins was determined using the Nanodrop, and purity estimated through SDS-PAGE.

Activity assessment using small molecules

25 2-3'-sialyllactose, 2-6'-sialyllactose, and colominic acid (2-8'; Sigma-Aldrich) were used as substrates to determine the hydrolytic specificity of the enzymes. Enzymes (0.05 µg) were mixed with a substrate (25 µM) in 20 mM Tris-HCl pH 6.8, and incubated for 30 minutes at 37°C, after which free sialic acids in the mixture were quantified, according to manufacturer's instructions (Sialic Acid Quantification Kit, Abcam).

30

Activity assessment using protein substrate

TNFR, EPO, Enbrel and fetuin (0.5 µg) were mixed together with varying concentrations of tested sialidases (1:40) or commercial sialidase (from NEB, according to manufacturer) and incubated for 30 minutes, after which the proteins were separated on 4-20% Novex gradient SDS-PAGE, and/or analyzed with SNA lectin blots.

5

Lectin blots

Separated proteins were transferred to PVDF membranes using the Trans-Blot Turbo Transfer System (BioRad). Membranes were blocked with lectin buffer, and subsequently incubated with the primary binder (SNA-biotin) and the secondary binder (HRP-streptavidin; VectorLabs) with washing steps in between. Chemi-luminescence was developed by the West Pico SuperSignal (ThermoFisher) kit, and detected in a ChemiDoc (BioRad).

Optimal enzymatic conditions

The sialidases were incubated with 2-3' sialyllactose (25 µM) in 20 mM Tris-HCl pH 8.0 to investigate the impact of NaCl (0-1.5 M) and ions (2 mM CaCl₂, 2 mM ZnCl₂, 5 mM EDTA). For pH optimum, the sialidases were incubated in acetic acid buffers (4.6 and 5.6), and Tris-HCl buffers (6.8, 7.4, 8.0, and 8.8). All samples were incubated for 15 minutes at room temperature after which the mixture was added 1:1 to a sialic acid quantification kit (Abcam), according to manufacturer's instructions. All values were expressed as relative activity in relation to the highest activity within each group.

Comparison with known sialidases from established biotechnology companies

A. muciniphila sialidases, sialidase mixtures, and bought sialidases from established brands (NEB P0743S, P0720S, and P0722S) were incubated with 2-3' and 2-6' sialyllactose, and colominic acid for 20 minutes in their respective optimal buffer together with the sialic acid quantification buffer (1:1 ratio), according to manufacturer's instructions.

30 **Results**

*All *A. muciniphila* sialidases express well and can be purified on His-columns*

Akkermansia annotated sialidases Am0705, Am0707, Am1758, and Am2058 were expressed recombinantly in *E. coli* and purified to high purity on His-columns (Fig. 1). The

proteins expressed well, with levels varying between 70-150 mg purified enzyme per L *E. coli* culture. Furthermore, the sialidases were highly soluble in Tris buffers, and could be concentrated to > 3 mg/mL without significant precipitation.

5 *Only three of the sialidases display activity towards a variety of protein substrates*

During initial screening of the purified sialidases towards different O- and N-linked glycoproteins, all sialidases except Am0705 displayed potent activity (data not shown). Due to inconsistent activity of Am2058, we continued the characterization of Am0707 and a mixture of Am0707/1757 (Mix, 1:1).

10

The sialidases have different specificities

Several sialidases display bond specificity, with limited ability to hydrolyze certain sialic acid bonds (e.g. 2-3, 2-6, and/or 2-8). To investigate the ability of the four *A. muciniphila* sialidases to act upon different bonds, we incubated the sialidases with specific substrates only having one type of the bonds present (2-3'-sialyllactose, 2-6'-sialyllactose, and colominic acid), and quantitated the free sialic acids (Fig. 2). Am1757 had a high specific activity against 2-3 bonds, while Am0707 had a broader, though lower, activity against all tested bonds. The combination of Am0707 and Am1757 (mix), resulted in a superior broad-spectrum product (Fig. 2).

20

The sialidases are NaCl sensitive and rely on divalent cations

To further investigate the conditions necessary for optimizing the sialidase activity, we investigated the dependence on ions, pH, and NaCl for the sialidases. The two sialidases behaved similarly, with a high sensitivity to EDTA and Zn^{2+} , while relying on Ca^{2+} . The enzymes had a higher activity in neutral to basic pH, and lost much of their activity in the presence of NaCl (Fig. 3).

25

A mixture of sialidases increases the overall efficiency of hydrolysis

Since the characterized sialidases had complementary activities, with Am1757 having a high 2-3 bond hydrolytic activity, and Am0707 also acting upon 2-6,8 bonds, we investigated if a mixture of the two enzymes could demonstrate high efficiency towards all the sialic acids bonds on native glycoproteins by varying the ratio of the two enzymes. A

30

mixture containing 1:40 Am1757 together with 1:40 Am0707 rapidly (<15 min) hydrolyzed all sialic acid bonds on fetuin (Fig. 4).

Benchmarking of the sialidase against small molecular substrates

5 In order to compare the efficiency of the sialidase mixture (GVS_Smix) to the single enzymes, as well as with brand competitors, we compared our enzymes with three commercially available sialidases from New England Biolabs (NEB). These were a broad spectrum sialidase from *Arthrobacter ureafaciens* (NEB A, the enzyme α 2-3,6,8,9 Neuraminidase A from *Arthrobacter ureafaciens*, catalog # P0722S; cleaves 2-3, 2-6, 2-8, 10 and 2-9 bonds), a narrow spectrum sialidase from *Streptococcus pneumoniae* (NEB S, the enzyme α 2-3 Neuraminidase S from *Streptococcus pneumoniae*, catalog # P0743S; cleaves 2-3 bonds only), and a general sialidase from *Clostridium perfringens* (NEB O, the enzyme α 2-3,6,8 Neuraminidase O from *Clostridium perfringens*, catalog # P0720S, cleaves 2-3, 2-6, 2-8 bonds). Enzymes were added as suggested by the manufacturers and 15 incubated with the substrate for 30 min at 37°C (Fig. 5).

As judged by an SDS-PAGE, the quantity of sialidase in the different samples is lower in the GVS_Smix than in the NEB, with Neuraminidase A possibly being the exception. Neuraminidase S displayed a limited ability to hydrolyze all the sialic acid bonds on the glycoprotein, while Neuraminidase A and the Neuraminidase O hydrolyzed 20 all sialic acids present on fetuin (Fig. 6). Likewise, as observed for activity against synthetic small molecule substrates, Smix hydrolyzed all sialic acid bonds on the glycoprotein, generating a narrow protein band, indicative of a complete asialylation. Am1757 treated samples, not able to hydrolyze 2-6 or 2-8 bonds, migrated as a higher molecular weight than those samples treated with Smix. The hydrolysis was confirmed 25 with SNA blotting, labeling sialic acids, suggesting that all broad-spectra sialidases efficiently removed all sialic acids present on fetuin, while 2-3 specific enzymes (*e.g.* neuraminidase S and Am1757) only removed a fraction of the sialic acids (Fig. 6).

Smix can efficiently release sialic acids from native proteins

30 Even though Smix could act on small semi-synthetic substrates with an efficiency similar to or better than the NEB products (Fig. 5), it was imperative to also investigate its activity against native proteins with different sialic acid bonds. As a model protein, fetuin

was chosen, since it has both 2-3 and 2-6 bonded sialic acids. Using SNA blotting, it became clear that the Smix at the chosen concentration could completely remove all sialic acids from fetuin, similarly to NEB A and O (Fig. 6). The two 2-3 specific enzymes Am1757 and NEB S could not, as expected, asialylate fetuin fully.

- 5 To more quantitatively determine the ability of Smix and Am1757 to release sialic acids from native proteins in comparison to the three NEB sialidases, we incubated different glycoprotein substrates (TNFaR, IgA, Plasminogen and Abatacept [Orencia]) with each sialidase for 15 + 15 minutes and quantified the released sialic acids. While certain substrates proved more difficult to hydrolyze than others, the GVS_Smix was at least comparable to the NEB products in all cases (Fig. 7) and demonstrated the most consistency, in that it showed high activity against all substrates. Although the individual enzymes of Am1757 and each of the NEB products had higher activities against one or another individual substrate, only the GVS_Smix showed consistently high activity against all substrates, making it a highly attractive tool for glycan analysis.
- 10
- 15

Example 2 – O-glycosidase

Materials and Methods

20 Expression and purification of an endo-O-glycosidase

- The *Streptococcus oralis* endo-N-acetyl-galactosaminidase was codon-optimized to express well in *E. coli* in the vector pET21a(+). The vector was transformed into BL21(DE3) Star cells. *E. coli* was routinely cultured in LB at 37°C, 200 rpm. In the presence of the plasmid, 100 µg/mL ampicillin was added. After overnight incubation, cultures were diluted 1:20 in fresh LB(amp), and grown until OD₆₂₀ ~ 0.7-0.8, after which recombinant protein expression was induced by addition of 1 mM IPTG, and the expression continued for 5 hours before the cells were collected and frozen. Frozen cells were thawed and resolved in His binding buffer (20 mM NaP pH 7.4, 500 mM NaCl, 20 mM imidazole), and sonicated for release of intracellular proteins. Cell debris was removed by centrifugation. Sterile filtered supernatant was affinity purified on a nickel column, and re-buffered in PBS on a PD-25 column. The concentration of the protein was
- 25
- 30

determined using the Nanodrop, and purity estimated through SDS-PAGE. The sequence of the protein is provided as SEQ ID NO: 3.

Activity assessment using small molecules

5 4-Methylumbelliferone (4MU) substrates and pNP-substrates of O-glycan core regions were used as substrates to determine the hydrolytic activity of the enzymes. Enzymes (1 µg) were mixed with a substrate (2 mM) and incubated for 15-120 minutes at 37°C, during which time fluorescence and absorbance (405 nm) was recorded, respectively.

10

Activity assessment using protein substrate

 TNFαR, EPO, Enbrel, fetuin, IgA, Orencia and plasminogen (0.5 µg) were mixed with O-glycosidase (1:40) either with or without the presence of Smix (1:40 + 1:40) or Am1757 (1:40) for 0-24 hours. The proteins were separated on 4-20% Novex gradient
15 SDS-PAGE.

Optimal enzymatic conditions

 The enzymes were incubated with their respective substrate (4MU or pNP) for investigation of NaCl (0-1.5 M) and ion (2 mM CaCl₂, 2 mM ZnCl₂, 5 mM EDTA)
20 dependency. For pH optimum, the enzymes were incubated in acetic acid buffers (50 mM pH 4.6 and 5.6), and Tris-HCl buffers (20 mM pH 6.8, 7.4, 8.0, and 8.8). All samples were incubated for 15 minutes at 37°C. All values were expressed as relative activity in relation to the highest activity within each group.

25 *Comparison with commercially available enzymes*

 The identified enzyme, as well as commercially available enzymes from established brands (NEB O-glycosidase from Enterococcus faecalis, catalog# P0733S, also as a bundle #E0540S) were incubated with their respective substrate(s) and optimized buffers, and incubated for 0-24 hours with different glycoproteins under native conditions
30 before being separated on SDS-PAGE.

Results

The Streptococcus oralis endo- α -N-acetyl-galactosaminidase expression/purification

During the discovery phase, two different endo- α -N-acetyl-galactosaminidases from *S. oralis* and *Bifidobacterium bifidum*, both being high molecular weight bacterial proteins (>200 kDa), were considered. The *B. bifidum* glycosidase was highly unstable, or
5 at least resulted in a high degree of fragmented parts after expression and affinity purification. While certain fragmentation also could be visualized for *S. oralis*, it was much less pronounced (Fig. 8A). Further, this enzyme was stable at 4°C for up to a week without any additional degradation (Fig. 8B). However, the expression level (ca 5-10 mg/L) is quite low. Further analysis suggested the fragmentation was due to the sonication.

10

The O-glycosidases can act upon synthetic pNP-labeled Core 1-3 O-glycans.

Continued analysis of the two O-glycosidases revealed a striking preference for Core 1 glycans, with a much lower activity towards core 2 and 3 (Fig. 9). Importantly, the glycosidase from *S. oralis* showed a significantly higher activity than the corresponding
15 gene from *B. bifidum*, so we decided to mainly focus on the *S. oralis* O-glycosidase.

Addition of MgCl₂ significantly increases the activity of the S. oralis O-glycosidase

In order to determine the optimal conditions for the O-glycosidase, the enzyme was incubated with the pNP Core 1 substrate at different conditions, under a range of pHs and
20 ions (Fig. 10). The O-glycosidase showed high activity at neutral to slightly basic pH, was completely inhibited by the presence of Zn²⁺, but had an increased activity in the presence of up to 8 mM MgCl₂ (Fig. 10C).

The S. oralis O-glycosidase can hydrolyze the glycans from native glycoproteins

25 To investigate the kinetics and doses necessary for mediating hydrolysis of all O-glycans, native TNF α R was incubated with varying amounts of O-glycosidase for 1-12 h in combination with the sialidase mixture characterized in Example 1 (1:1 mix of Am1757:Am0707), or with sialdase mixture only as . As shown in Figure 11, even comparably low concentrations of the enzyme (e.g. ca 0.1 μ g, 1:10) could hydrolyze the
30 substrate within one hour. Increasing the incubation time to 12 hours resulted in an ability of the enzyme to fully hydrolyze the substrate at a ratio of 1:50 with a high concentration

of sialidase. Of importance to note, the enzyme is, in some aspects, a “nothing or all” enzyme, barely resulting in semi-hydrolyzed glycoproteins, but rather non-hydrolyzed or completely hydrolyzed proteins.

5 *S. oralis* O-glycosidases can act upon different native glycoproteins

To further investigate whether the O-glycosidase only acts upon TNF α R or can act on several native glycoproteins, we incubated seven different glycoproteins with combinations of hydrolytic enzymes (Fig. 12). The *S. oralis* O-glycosidase, as well as the *B. bifidum* enzyme (in combination with sialidases) were able to hydrolyze all native
10 proteins after an overnight incubation, even though the activity on plasminogen was difficult to assess due to the high molecular weight of the protein. Both enzymes did however strongly depend on the presence of sialidases since terminal sialic acids inhibited the activity.

15 *The S. oralis* O-glycosidase is superior in hydrolyzing native glycoproteins as compared to the NEB product portfolio.

To compare the activity of the *S. oralis* glycosidase against existing commercial products, we compared their ability to hydrolyze the TNF α R using varying amounts of glycosidases for either 1 h or 16 hours. Using a ca 1:5 enzyme:substrate ratio, the *S. oralis*
20 glycosidase was able to fully hydrolyze its substrate in 1 h. Even at a 1:1 enzyme:substrate ratio, the NEB O-glycosidase did not hydrolyze all O-linked glycans, but only acted upon a few easily accessible glycans. A further incubation (e.g. 16 h) allowed for a 1:50 enzyme:substrate ratio while still maintaining full effect of the *S. oralis* glycosidase. However, the NEB O-glycosidase still failed to fully deglycosylate the product, indicating
25 that denaturation is pivotal for its function, while the *S. oralis* O-glycosidase product has a high activity also against native proteins (Fig. 13). Not only can the *S. oralis* O-glycosidase act upon native proteins, but is also able to hydrolyze glycans from denatured proteins (data not shown).

30 *Glycan composition influence the activity of the O-glycosidase*

To further study the necessity of specific glycans for O-glycosidase activity, TNF α R was pre-incubated with different enzymes to remove individual glycans before

addition of the O-glycosidase. The removal of the terminal sialic acids was critical for activity, as was the presence of galactoses, indicating that the O-glycosidase can not remove single (terminal) GalNAcs (Fig. 14).

5 *O-glycosidase is highly efficient in removing O-glycans from native proteins*

Once the final concentration of the combined O-glyk + Smix composition was determined (1:40), we repeated the comparison with the competitor brand product (NEB). While some hydrolysis can be detected on the TNF α R using the NEB products, it is evident that the hydrolysis is not complete, not even after 12 h. Opposite to this, the
10 combined O-glyk + Smix composition resulted in a full hydrolysis of the glycoprotein within 4 hours, supported by the lectin blot (Fig. 15).

Similarly, Enbrel showed an identical pattern, with the combined O-glyk + Smix composition able to fully hydrolyze it, while the NEB product did not. However, for feutrin, both products were seemingly equally efficient (Fig. 16).

15

Both 2-3 and 2-6 linked sialic acids inhibits the activity of the O-glycosidase

In order to evaluate the mutual effect of 2-3 or 2-6 linked sialic acids for O-glycosidase activity, we incubated the O-glycosidase with both 2-3 specific, or broad-spectrum sialidases (e.g. Am1757 and Smix, respectively). Though preliminary data
20 suggest that Am1757 and Am0707 can release equal quantities of sialic acids from glycoproteins, treatment of the glycoprotein with Am1757 (or Smix) resulted in a faster hydrolysis of the substrate (TNFaR). However, for full hydrolysis it was critical to treat the glycoprotein with a broad-spectrum sialidase, to also remove 2-6 (or 2-8) sialic acids (Fig. 17).

25

The native activity of the GVS O-glycosidase bundle is due to highly efficient sialidase activity

To determine the impact of the individual components in the combined O-glyk + Smix composition, O-glyk and NEB were incubated with four different sialidase products
30 (GVS_Smix, NEB A, NEB S, and NEB O). The combined O-glyk + Smix composition efficiently hydrolyzed all O-glycans from Enbrel (Fig. 18), while changing the sialidase to any of the NEB products resulted in a significantly lowered activity of O-glyk, with the

NEB sialidase A being the most potent. However, while the NEB O-glycosidase bundle (NEB O-glycosidase with NEB sialidase O) could not hydrolyze the Enbrel or TNF α R (Fig. 15, 16), by using GVS_Smix in place of the NEB sialidase O, full hydrolysis of the glycoprotein was observed, suggesting that the efficiency of the combined O-glyk + Smix composition relies at least in part on the ability of the sialidases to hydrolyze all 2-3 and 2-6 bonds on the glycoprotein.

Example 3 – O-glycoprotein specific endoprotease

We recently determined that endoprotease activity of the polypeptide consisting of SEQ ID NO: 12 at least in part relied on specific sialic acid bonds, necessitating the removal of both 2-3 and 2-6 linked sialic acids for full effect. To determine the individual role of specific sialic acid bonds for the endoprotease activity, we incubated Etanercept with different sialidases in combination with the endoprotease for 30 min – 20 h. Removal of 2-3 bonds seemed sufficient to enhance endoprotease activity (Fig. 19). Sialidases used were (1) Am0707 = polypeptide of SEQ ID NO: 3; (2) Am1757 = polypeptide of SEQ ID NO: 6; Mix = 1:1 combination of (1) and (2).

Sequences

SEQ ID NO: 1 – sialidase, Am0707 – wildtype (signal sequence underlined)
MTWLLCGRGKWNKVKRMNSVFKCLMSAVCAVALPAFGQEEKTGFPDRAVTVFSAGEGNPYASIRIPALLSIGKQLLAFAEGRYKNTDQGENDIIMSVSKNGGKTWSRPAIAKAHGA
TGFNNPCPVYDAKTRTVTVVFQRYPA
GKERQPNIPDGWDDEKCI
RNFMIQSRNGGSSWTKPQEI
TKTTKRPSGVDIMASGPNAGTQLKSGAHKGR
LVI
PMNEGPF
GKQV
AMSYDNGKTWPVKKLLGEGGFAYSS
LAMVEPGIVGVLYEENQEH
IKKLFVPITMEWLT
DGEDTGLAPGKKAPVLK

SEQ ID NO: 2 – sialidase, Am0707 – wildtype without signal
QEEKTGFPDRAVTVFSAGEGNPYASIRIPALLSIGKQLLAFAEGRYKNTDQGENDIIMSVSKNGGKTWSR
RAIAKAHGA
TGFNNPCPVYDAKTRTVTVVFQRYPA
GKERQPNIPDGWDDEKCI
RNFMIQSRNGGSSWTKPQEI
TKTTKRPSGVDIMASGPNAGTQLKSGAHKGR
LVI
PMNEGPF
GKQV
AMSYDNGKTWPVKKLLGEGGFAYSS
LAMVEPGIVGVLYEENQEH
IKKLFVPITMEWLT
DGEDTGLAPGKKAPVLK

SEQ ID NO: 3 – sialidase, Am0707 – with additional N terminal methionine and C terminal linker+His6 tag (bold & underlined)
MQEEKTGFPDRAVTVFSAGEGNPYASIRIPALLSIGKQLLAFAEGRYKNTDQGENDIIMSVSKNGGKTWSR
PRAIAKAHGA
TGFNNPCPVYDAKTRTVTVVFQRYPA
GKERQPNIPDGWDDEKCI
RNFMIQSRNGGSSWTKPQEI
TKTTKRPSGVDIMASGPNAGTQLKSGAHKGR
LVI
PMNEGPF
GKQV
AMSYDNGKTWPVKKLLGEGGFAYSS
LAMVEPGIVGVLYEENQEH
IKKLFVPITMEWLT
DGEDTGLAPGKKAPVLK**GSGLEHHHHHH**

SEQ ID NO: 4 - sialidase, Aml757 - wildtype (signal sequence underlined)
 MKNLLFALLTGSFCCCYAQQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGRKEPDYLVSRDNGKTWE
 MKAAPAGYPPNYGGIPKESPAIVRNPLTREFIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKN
 5 LKKLGGIMRTPVFFVNKRRVIVPFHNMGGGTFKHISDDGGLTWHVSRNGVTSRHEARPPHQGVRFNNAVEA
 TVLEMKDGTWLALARTSQDQAWQAFSKDYGETWSKPEPSRFFGTLTMTNLGRLLDGTIVSLWTNTMALPENAT
 AGNGTWEDVFTNRDSSHIAMSGDEGKTWYGFREIILDEHRNHPGYATLDGPEDRGRKHQSEMVQLDKNRILISL
 GQHKNHRRLLVIVDRRWGAKTRATQTGKDLDSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSGGTKKVLQIKR
 10 LDDPELVNEKSNVDYRNGGATWNFPNGTTGLVKFRFRVVDGEQADDSGLQVSLTDRLFNACDSTTKDYALFTF
 PIRLKPAPHLLLGMMKVPFPTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANKSPNGASYIHFIISTGSQPDAG
 ILDDTVNARVK

SEQ ID NO: 5 - sialidase, Aml757 - wildtype without signal
 QQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGRKEPDYLVSRDNGKTWEMKAAPAGYPPNYGGIPKE
 15 SPAIVRNPLTREFIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKNLKKLGGIMRTPVFFVNKGR
 RVIVPFHNMGGGTFKHISDDGGLTWHVSRNGVTSRHEARPPHQGVRFNNAVEATVLEMKDGTWLALARTSQ
 DQAWQAFSKDYGETWSKPEPSRFFGTLTMTNLGRLLDGTIVSLWTNTMALPENATAGNGTWEDVFTNRDSSHIA
 MSGDEGKTWYGFREIILDEHRNHPGYATLDGPEDRGRKHQSEMVQLDKNRILISLQHKNHRRLLVIVDRRWG
 AKTRATQTGKDLDSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSGGTKKVLQIKRLDDPELVNEKSNVDYRNG
 20 GATWNFPNGTTGLVKFRFRVVDGEQADDSGLQVSLTDRLFNACDSTTKDYALFTFPIRLKPAPHLLLGMMKVP
 FPTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANKSPNGASYIHFIISTGSQPDAGILLDDTVNARVK

SEQ ID NO: 6 - sialidase, Aml757 - with additional N terminal methionine
 and C terminal linker+His6 tag (bold & underlined)
 25 **MQQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGRKEPDYLVSRDNGKTWEMKAAPAGYPPNYGGIPK**
ESPAIVRNPLTREFIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKNLKKLGGIMRTPVFFVNKGR
RRVIVPFHNMGGGTFKHISDDGGLTWHVSRNGVTSRHEARPPHQGVRFNNAVEATVLEMKDGTWLALARTS
QDQAWQAFSKDYGETWSKPEPSRFFGTLTMTNLGRLLDGTIVSLWTNTMALPENATAGNGTWEDVFTNRDSSHIA
MSGDEGKTWYGFREIILDEHRNHPGYATLDGPEDRGRKHQSEMVQLDKNRILISLQHKNHRRLLVIVDRRWG
AKTRATQTGKDLDSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSGGTKKVLQIKRLDDPELVNEKSNVDYRNG
GGATWNFPNGTTGLVKFRFRVVDGEQADDSGLQVSLTDRLFNACDSTTKDYALFTFPIRLKPAPHLLLGMMKVP
PFTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANKSPNGASYIHFIISTGSQPDAGILLDDTVNARVK
GGSGLEH
HHHHH

35 SEQ ID NO: 7 - O-glycosidase from *S. oralis*
 Wildtype (signal sequence underlined; C terminal element of LPXTG cell
 wall anchor motif bold underlined)
 MDKRFFEKRCKFSIRKFTLGVASVMIGATFFAASPVADQARVGSTDNLPSELADLDKKASDEGHDFDKEAAA
 40 QNPGSAETTEGPQTEEELLAQEKEKSEKPSNLPKELEDKLEKAEDNGREVDKDQLAQDTGKLPEDVAKTTNG
 ELNYGATVGIKTPSGEGSGIVVAKDLVLTVSHNFIKDSQEGNIRKVVNDNQDGDGIYSISYPLGPDVKFSSKD
 IIHWDRGYLKGFKNDLALVRLRTVLENTPVEVTKKPVVKKIGDKLHVFGYPEGKLNPIVNTTVDFAEPYGEG
 VQGIGYQGGKPGASGGGIFDTEGKLVGVHQNGVVGKRSGGILFSPAQLKWIQDHMQGIISSVKPADLEEKEKPA
 EEKPKEDKPAAAKPETPKAVTPEWQTVANKEQQGTVTIREEKGVRYNQLSSTAQNNDNGKPALFEKQGLTVDA
 45 NGNATVDLTFKDDSEKKSREFGVFLFKFDTKNNVFGVDQGGWFWEYKTPGNSTWYKGNRVAAPFPGSVNRLS
 ITLKSQGQLNASNNDVNLFDTVTLPGAVNENLNKNEKILLKAGTYSNDRTVVSVKTDNQEYKADDTPAQKET
 GPAVDDSKVTYDTIQSKVLKAVIDQAFPRVKEYTLNGHTLPQGVQQFNQVFINNHRITPEVTYKINETTAEY
 LMKLRDDAHLINAEMTVRLQVVDNQLHFDVTKIVNHNQVTPGQKIDDERKLLSTISFLGNALVSVSSDQAGAK
 FDGATMSNNTHVSGDDHIDVTNPMKDLAKGYMGFVSTDKLAAGVWSNSQNSYGGGSNDWTRLTAYKETVGNA
 50 NYVGIHSSEWQWEKAYKGI VFP EYTKELPSAKVVITEDANADNKVDWQDGAIAYSIMNPNQGWKVKDITAY
 RIAMNFGSQAQNPFLMTLDGIKKINLHTDGLQGQVLLKGYGSEGHDSGHLNYADIGKRIGGVDFKTLIEKAK
 KYGAHLGIHVNASETYPESKYFNENILRKNPDGSYSYGNWLDQGINIDAAYDLAHGRLARWEDLKKKLGEGL
 DFIYVDVWGNQSGDNGAWATHVLAKEINKQGWRF AIEWGHGGEYDSTFQHWAADLT YGGYTNGKINSAITRF
 IRNHQKDSWVGDRSYGGAANYPLGGYSMKDFEGWQGRSDYNGYVTNLFAHDVMTKYFQHFTVSKWENGTPV
 55 TMTDNGSTYKWTPEMKVELVDAAGNKVVVTRKSNVDNSPQYRERTVTLNGRVIQDSAYLTPWNWDANGKKLP
 TEKEKMYFFNTQAGATTWTLPSDWANSKVYLYKLTDQGKTEEQELTVDGKITLDDL ANQPYVLYRSKQTNPE
 MSWSEGMMHIYDQGFNSGTLKHWITISGDASKAEIVKSQGANEMLRIOGNKSKVSLTQKLTGLKPNTKYAVVYGV
 DNRNNAKASITVNTGEKEVTYTNKSLALNYIKAYAHNNRRENATVDDTSYFQNMAYFFTGTGSDVSNVTLLS

REAGDEATYFDEIRTFENNSSMYGDKHDTGQGTFFKQDFENVAQGIFFPVVGGVEGVEDNRTHLSEKHDPYTQR
 GWNGKKVDDVIEGNWSLKTNGLVSRRLVYQTI PQNFRFEAGKTYRVTFEYEAGSDNTYAFVVGKGEFQSGRR
 GTQASNLEMHELPTNTWTDSSKAKKVTFVLTGAETGDTWVGIIYSTGNASNTRGDAGGNANFRGYNDFMMDNLQI
 EEITLTGKMLTENALKNYLPTVAMTNYTKESMDALKEAVFNLSQADDDISVEEARAEIAKIEALKNALVQKKT
 5 ALVAEDFESLDAPAQPGEGLNADFNGVSSLWHTSWNGGDVGKPATMVLKEPTEITGLRYVPRASDSNGNLRD
 VKLVVTDESGKEHTFNVTDWPNNNKPKDIDFGKTIKAKKIVLTGKTYGDGGDKYQSAELIFTRPQVAETPL
 DLSGYEAALAKAQKLTDKDNQEEVASVQASMKYATDNHLLTERMVAYFADYLNQLKDSATKPDAPTSSKGEEQ
 PPVLDVPEFKGGVNATEAAVHEVPEFKGGVNAVQALVHELPEYKGGANAVLAAANEVPEYKGGANAVEALVNE
 KPAYTGVLATAGDQAAPTVEKPEYPLTPSPVADTKTPGAKDEEKLPATGEHSSEVALFLASVSIALSAAVLAT
 10 KRKEE

SEQ ID NO: 8 - O-glycosidase from *S. oralis*
 Wildtype with signal sequence removed (C terminal element of LPXTG cell
 wall anchor motif bold underlined)

DQARVGSTDNLPSSELADLDKKASDEGHDFDKEAAAQNPGSAETTEGPQTEEEELLAQEKEKSEKPSNLPKELED
 KLEKAEDNGREVDKDLAQDTGKLVPEVDAKTTNGELNYGATVKIKTPSGEGSGIVVAKDLVLTVSHNFIKDS
 QEGNIRKVVNDNQGDGDIYSISYPGLPDVKFSKKDIHWDREGYLGKFNLDLALVRLRTVLENTPVEVTKKPV
 VKKIGDKLHVFGYPEGKLNPIVNTTVDFAEPYGEGVQIGYQGGKPGASGGGIFDTEGKLVGVHONGVVGKRS
 20 GGILFSPAQLKWIQDHMQGISSVKPADLEEKEKPAEEKPKEDKPAAPKAVTPKAVTPEWQTVANKEQQGTVTI
 REEKGVRYNQLSSTAQNNDNGKPALEFEKQGLTVDANGNATVDLTFKDDSEKGSRFVFLKFKDTKNNVFGVY
 DQGGWFWEYKTPGNSTWYKGNRVAAPPGSVNRLSITLKSQGLNASNNDVNLFDVTPLPGAVNENLKNKKI
 LLKAGTYSNDRTVVSVKTDNQEGVKADDTPAQKETGPAVDDSKVTYDTIQSKVLKAVIDQAFPRVKEYTLNGH
 TLPQGVQGFQNFVFINNHRTPEVTYKKINETTAEYLMKLRDDAHLINAEMTVRLQVVDNQLHFDVTIKIVNHNQ
 25 VTPGQKIDDERKLLSTISFLGNALVSVDQAGAKFDGATMSNNTHVSGDDHIDVTNPMKDLAKGYMGFVST
 DKLAAGVWSNSQNSYGGGSNDWTRLTAYKETVGNANYVGIHSSEWQWEKAYKGI VPEYTKELPSAKVVITED
 ANADNKVDWQGAIAYSIMNPNQGWKVKDITAYRIAMNFGSQAQNPFLMTLDGKIKINLHDTGLGQGVLLK
 GYGSEGHDSGHLNYADIGKRIGGVDFKTLIEKAKKYGAHLGIHVNASETYPESKYFNENILRKNPDGYSYSY
 WNWLDQGINIDAAAYDLAHGRLARWEDLKKKLGEGLDFIYVDVWGNQSGDNGAWATHVLAKEINKQGRFAIE
 30 WGHGGEYDSTFQHWAADLTYYGYTNKGINSATIRFIRNHQKDSWVGDIRSYGGAANYPLGGYSMKDFEGWQG
 RSDYNGYVTNLFAHDMTKYFQHFTVSKWENGTPVTMTDNGSTYKWTPEMKVELVDAAGNKVVTRKSNVNS
 PQYRERTVTNLGRVIQDGSAYLTPWNWDANGKKLPTEKEKMYFNTQAGATTWTLPDWSANKVYLYKLTDQG
 KTEEQELTVDGKITLDDLANQPYVLYRSKQTNPEMSWSEGMHIYDQGFNSGTLKHWTISGDASKAEIVKSQG
 ANEMLRIQGNKSKVSLTQKLTGLKPNTKYAVYVGVNDNRNNAKASITVNTGEKEVTYTNKSLALNYIKAYAHN
 35 NRRENATVDDTSYFQNMAYFTTGSVSNVTLTSLREAGDEATYFDEIRTFENNSSMYGDKHDTGQGTFFKQDF
 ENVAQGIFFPVVGGVEGVEDNRTHLSEKHDPYTQRGWNGKKVDDVIEGNWSLKTNGLVSRRLVYQTI PQNFR
 FEAGKTYRVTFEYEAGSDNTYAFVVGKGEFQSGRRGTQASNLEMHELPTNTWTDSSKAKKVTFVLTGAETGDTW
 VGIYSTGNASNTRGDAGGNANFRGYNDFMMDNLQIEEITLTGKMLTENALKNYLPTVAMTNYTKESMDALKEA
 VFNLSQADDDISVEEARAEIAKIEALKNALVQKKTALVAEDFESLDAPAQPGEGLNADFNGVSSLWHTSWNG
 40 GDVGKPATMVLKEPTEITGLRYVPRASDSNGNLRDVKLVVTDESGKEHTFNVTDWPNNNKPKDIDFGKTIKAK
 KIVLTGKTYGDGGDKYQSAELIFTRPQVAETPLDLSGYEAALAKAQKLTDKDNQEEVASVQASMKYATDNH
 LLTERMVAYFADYLNQLKDSATKPDAPTSSKGEEQPPVLDVPEFKGGVNATEAAVHEVPEFKGGVNAVQALVH
 ELPEYKGGANAVLAAANEVPEYKGGANAVEALVNEKPAYTGVLATAGDQAAPTVEKPEYPLTPSPVADTKTPG
 AKDEEKLPATGEHSSEVALFLASVSIALSAAVLATKRKEE
 45

SEQ ID NO: 9 - O-glycosidase from *S. oralis*
 Wildtype with signal sequence and C terminal element of LPXTG cell
 wall anchor motif both removed

DQARVGSTDNLPSSELADLDKKASDEGHDFDKEAAAQNPGSAETTEGPQTEEEELLAQEKEKSEKPSNLPKELED
 KLEKAEDNGREVDKDLAQDTGKLVPEVDAKTTNGELNYGATVKIKTPSGEGSGIVVAKDLVLTVSHNFIKDS
 QEGNIRKVVNDNQGDGDIYSISYPGLPDVKFSKKDIHWDREGYLGKFNLDLALVRLRTVLENTPVEVTKKPV
 VKKIGDKLHVFGYPEGKLNPIVNTTVDFAEPYGEGVQIGYQGGKPGASGGGIFDTEGKLVGVHONGVVGKRS
 50 GGILFSPAQLKWIQDHMQGISSVKPADLEEKEKPAEEKPKEDKPAAPKAVTPKAVTPEWQTVANKEQQGTVTI
 REEKGVRYNQLSSTAQNNDNGKPALEFEKQGLTVDANGNATVDLTFKDDSEKGSRFVFLKFKDTKNNVFGVY
 DQGGWFWEYKTPGNSTWYKGNRVAAPPGSVNRLSITLKSQGLNASNNDVNLFDVTPLPGAVNENLKNKKI
 LLKAGTYSNDRTVVSVKTDNQEGVKADDTPAQKETGPAVDDSKVTYDTIQSKVLKAVIDQAFPRVKEYTLNGH
 55 TLPQGVQGFQNFVFINNHRTPEVTYKKINETTAEYLMKLRDDAHLINAEMTVRLQVVDNQLHFDVTIKIVNHNQ

VTPGQKIDDERKLLSTISFLGNALVSVSSDQAGAKFDGATMSNNTHVSGDDHIDVTNPMKDLAKGYMYGFVST
 DKLAAGVWSNSQNSYGGGSNDWTRLTAYKETVGNANYVGIHSSEWQWEKAYKGIVFPEYTKELPSAKVVITED
 ANADNKVDWQDGAIAYRSIMNPNQGWKVKDITAYRIAMNFGSQAQNPFLMTLDGIKKINLHTDGLGQGVLK
 GYGSEGHDSGHLNYADIGKRIGGVDFKTLIEKAKKYGAHLGIHVNASETYEPESKYFNENILRKNPDGSYSY
 5 WNWLDQGINIDAAYDLAHLARWEDLKKKLGEGLDFIYVDVWNGQSGDNGAWATHVLAKINKQGWRF
 WGHGGEYDSTFQHWAADLTYYGGYTNKGINSATIRFIRNHQKDSWVG DYRSYGGGAANYPLGGYSMKDFEGWQ
 RSDYNGYVTNLFAHDVMTKYFQHFTVSKWENGTPVTMTDNGSTYKWTPEMKVELVDAAGNKVVVTRKSN
 PQRERTVTNLGRVIQDGSAYLTPWNWDANGKKLPTEKEKMYFNTQAGATTWTLPSDWANSKVLYKLTDQ
 10 KTEEQELTVDGKITLDLLANQPYVLYRSKQTNPEMSWSEGMHIYDQGFNSGTLKHWTISGDASKAEIVKSQ
 ANEMLRIQGNKSKVSLTQKLTGLKPNTRYAVYVGVNDRSNAKASITVNTGEKEVTYTNKSLALNYIKAYAH
 NNRRENATVDDTSYFQNMAYFFTTGSDVSNVTLLTSREAGDEATYFDEIRTFENSSMYGDKHDTGQGT
 ENVAQGIFFVVGVEGVEDNRTHLSEKHDPYTQRGWNGKKVDDVIEGNWSLKTNGLVSRRLVYQTI
 FEAGKTYRVTFEYEAGSDNTYAFVVGKGEFQSGRRGTQASNLEMHELPTWTDSKKAKKVTFLVTGAETGDT
 VGIYSTGNASNTRGDAGGNANFRGYNDFMMDNLQIEEITLTGKMLTENALKNYLPTVAMTNYTKESMDALKEA
 15 VFNLSQADDDISVEEARAEIAKIEALKNALVQKKTALVAEDFESLDAPAQPGELENAFDGNVSSLWHTSWNG
 GDVGKPATMVLKEPTEITGLRYVPRASDSNGNLRDVKLVVTDESKEHTFNVTDWPNNNPKPIDDFGKT
 KIVLTGKTGYDGGDKYQSAEELIFTRPQVAETPLDLSGYEAALAKAQKLTDKDNQEEVASVQASMKYATDNH
 LLTERMVAYFADYLNQLKDSATKPDAPTSKSGEEQPPVLDVPEFKGGVNATEAAVHEVPEFKGGVNAVQALVH
 ELPEYKGGANAVLAAANEVPEYKGGANAVEALVNEKPAYTGVLATAGDQAAPTVEKPEYPLTPSPVADTKTPG
 20 AKDEEKLPA

SEQ ID NO: 10 - O-glycosidase from *S. oralis* with additional N terminal
 Met, C terminal GSGLE-His6tag (bold underlined), and signal sequence and
 25 C terminal element of LPXTG cell wall anchor motif removed

MDQARVGSTDNLPSELADLDKKASDEGHDFDKEAAAQNPGSAAETTEGPQTEEEELLAQEKEKSEKPSNLPKELE
 DKLEKAEDNGREVDDKDLAQDTGKLVPEVAKTTNGELNYGATVKIKTPSGEGSGIVVAKDLVTVSHNFIKD
 SQEGNIRKVVNDQGDGDIYSISYPGLPDVKFSKKDIHWDREGYLGKFNKDLALVRLRVTLENTPVEVTKKP
 30 VVKKIGDKLHVFGYPEGKLNPIVNTTVDFAEPYGEGVQGGIGYQGGKPGASGGGIFDTEGKLVGVHONGVVGKR
 SGGILFSPAQLKIWDHMQGISSVKPADLEEKEKPAEEKPKEDKPAAPKPKAVTPEWQTVANKEQQGTVT
 IREEKGVRYNQLSSTAQNNDNGKPALEKQGLTVDANGNATVDLTFKDDSEKGSRFVFLFKDTKNNVFG
 YDQGGWFWEYKTPGNSTWYKGNRVAAPPEGSVNRLSITLKSDGQLNASNNDVNLFDVTVTPGAVNENLNKNEK
 ILLKAGTYSNDRTVSVKTDNQEGVKADDTPAQKETGPAVDDSKVTYDTIQSKVLKAVIDQAFPRVKEYTLNG
 35 HTLPQVQVQNFQVFINNHRTPEVTYKKINETTAEYLMKLRDDAHLINAEMTVRLQVVDNQLHFDVTKIVNH
 QVTPGQKIDDERKLLSTISFLGNALVSVSSDQAGAKFDGATMSNNTHVSGDDHIDVTNPMKDLAKGYMYGFVS
 TDKLAAGVWSNSQNSYGGGSNDWTRLTAYKETVGNANYVGIHSSEWQWEKAYKGIVFPEYTKELPSAKVVITE
 DANADNKVDWQDGAIAYRSIMNPNQGWKVKDITAYRIAMNFGSQAQNPFLMTLDGIKKINLHTDGLGQGVLL
 KGYGSEGHDSGHLNYADIGKRIGGVDFKTLIEKAKKYGAHLGIHVNASETYEPESKYFNENILRKNPDGSYSY
 40 GWNWLDQGINIDAAYDLAHLARWEDLKKKLGEGLDFIYVDVWNGQSGDNGAWATHVLAKINKQGWRF
 EWHGGEYDSTFQHWAADLTYYGGYTNKGINSATIRFIRNHQKDSWVG DYRSYGGGAANYPLGGYSMKDFEGWQ
 GRSDYNGYVTNLFAHDVMTKYFQHFTVSKWENGTPVTMTDNGSTYKWTPEMKVELVDAAGNKVVVTRKSN
 SPQRERTVTNLGRVIQDGSAYLTPWNWDANGKKLPTEKEKMYFNTQAGATTWTLPSDWANSKVLYKLTDQ
 GKTEEQELTVDGKITLDLLANQPYVLYRSKQTNPEMSWSEGMHIYDQGFNSGTLKHWTISGDASKAEIVKSQ
 45 GANEMLRQGNKSKVSLTQKLTGLKPNTRYAVYVGVNDRSNAKASITVNTGEKEVTYTNKSLALNYIKAYAH
 NNRRENATVDDTSYFQNMAYFFTTGSDVSNVTLLTSREAGDEATYFDEIRTFENSSMYGDKHDTGQGT
 FENVAQGIFFVVGVEGVEDNRTHLSEKHDPYTQRGWNGKKVDDVIEGNWSLKTNGLVSRRLVYQTI
 RFEAGKTYRVTFEYEAGSDNTYAFVVGKGEFQSGRRGTQASNLEMHELPTWTDSKKAKKVTFLVTGAETGDT
 WGIYSTGNASNTRGDAGGNANFRGYNDFMMDNLQIEEITLTGKMLTENALKNYLPTVAMTNYTKESMDALKEA
 50 AVFNLSQADDDISVEEARAEIAKIEALKNALVQKKTALVAEDFESLDAPAQPGELENAFDGNVSSLWHTSWN
 GGDVGKPATMVLKEPTEITGLRYVPRASDSNGNLRDVKLVVTDESKEHTFNVTDWPNNNPKPIDDFGKT
 KIVLTGKTGYDGGDKYQSAEELIFTRPQVAETPLDLSGYEAALAKAQKLTDKDNQEEVASVQASMKYATDNH
 HLLTERMVAYFADYLNQLKDSATKPDAPTSKSGEEQPPVLDVPEFKGGVNATEAAVHEVPEFKGGVNAVQALV
 HELPEYKGGANAVLAAANEVPEYKGGANAVEALVNEKPAYTGVLATAGDQAAPTVEKPEYPLTPSPVADTKTP
 55 GAKDEEKLPA**GSGLEHHHHHH**

SEQ ID NO: 11 - O-glycoprotein-specific endoprotease

5 EVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGLDIKS
GRVNIIEYKAKNPAAHYPYENGGGWKAAQELDEFFKAHPDRKKSQHTLIIMPTWNEKNGPDNPGGVPFYGMG
RNCFALDYPAFDIKHLGQKTREGRLLTKWYGGMAHELGHGLNLPNHQTASDGKKYGTALMGSGNYTFTGTSPT
FLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQ
DYDAVSFSRRLGKKS GKFSMKIDKKELEGLNNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKS

10 SEQ ID NO: 12 - O-glycoprotein-specific endoprotease (LS)
MEVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGLDIKS
PGRVNIIEYKAKNPAAHYPYENGGGWKAAQELDEFFKAHPDRKKSQHTLIIMPTWNEKNGPDNPGGVPFYGM
GRNCFALDYPAFDIKHLGQKTREGRLLTKWYGGMAHELGHGLNLPNHQTASDGKKYGTALMGSGNYTFTGTS
TFLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVN
QDYDAVSFSRRLGKKS GKFSMKIDKKELEGLNNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKSGSG
HHHHHH

15 SEQ ID NO: 13 - encodes SEQ ID NO: 10
ATGGACCAAGCGCGTGTGGGTAGCACCGATAACCTGCCGAGCGAGCTGGCGGATCTGGACAAGAAAGCGAGCG
ACGAAGGCCACGATTTTGAAGAGAGAGCGCGCGCGAGAACCCGGGTAGCGCGGAAACCACCGAAGGTCCGCA
GACCGAGGAAGAGCTGCTGGCGCAAGAAAAAGAGAAGAGCGAGAAGCCGAGCAACCTGCCGAAAGAACTGGAG
GATAAACTGGAAAAGGCGGAGGACAACGGTCTGTAAGTGGATAAAGACCAGCTGGCGCAAGACACCGGCAAGC
20 TGGTGCCGAGGATGTTGCGAAAACCACCAACGGTGAAGTGAAGTACGGCGCGACCGTTAAAAATTAGACCCC
GAGCGGCGAGGGTAGCGGTATTGTGGTTGCGAAGGACCTGGTGCTGACCGTTAGCCACAACCTTCATTAAGGAT
AGCCAGGAAGGTAATATCCGTAAAGTGGTTGATAACGACCAAGGCGATGGTGACATCTACAGCATTAGCTATC
CGGGCCTGCCGACGTTAAGTTCAGCAAGAAAAGATATCATCCACTGGGACCGTGAGGGTTACCTGAAAAGGCTT
CAAGAACGATCTGGCGCTGGTGCGTCTGCGTACCGTTCTGGAAAACACCCCGGTTGAGGTGACCAAGAAACCG
25 GTGGTTAAGAAAATTGGTGACAAGCTGCACGTGTTTGGTTATCCGGAGGGCAAACCTGAACCCGATCGTGAACA
CCACCGTTGATTTTCGCGGAACCGTACGGCGAGGGTGTTCAGGGCATTGGTTATCAAGGTGGCAAACCGGGCGC
GAGCGGTGGCGGTATCTTTGACACCGAAGGCAAGCTGGTTGGCGTGCACCAGAACGGTGTGGTTGGCAAACGT
AGCGGCGGTATTCTGTTACGCGCGGCAACTGAAGTGGATTACAGGACCATGCAAGGTATCAGCAGCGTGA
AACCAGCGGATCTGGAAGAGAAAAGAGAAGCCGCGCGGAAGAGAAAACCGAAGGAAGACAAGCCGCGCGCGGAA
30 GCCGGAACCCCGAAAGCGGTTACCCCGAGTGCGAAACCGTGGCGAACAAGGAACAGCAAGGTACCGTTACC
ATCCGTGAAGAGAAAGGCGTGCGTTACAACAGCTGAGCAGCACCGCGCAAAACGATAACGACGGCAAGCCGG
CGCTGTTTGAAGAACAGGGTCTGACCGTTGACGCGAACGGCAACGCGACCGTGGATCTGACCTTCAAGGACGA
TAGCGAAAAGGCAAGAGCCGTTTTCGGCGTTCCTGAAATTCAAGGACACCAAAAACAACGTTTTTGTGGGT
TACGATCAAGGCGGTTGGTTCTGGGAGTATAAGACCCCGGTAACAGCACCTGGTACAAGGGTAACCGTGTGG
35 CGCGCGGAAACCGGTAGCGTGAACCGTCTGAGCATTACCCTGAAAAGCGACGCGCAGCTGAACCGTGACAA
CAACGATGTGAACCTGTTGACACCGTTACCCTGCCGGGTGCGGTGAACGAAAACCTGAAGAACGAGAAGAAA
ATCCTGCTGAAAGCGGGCACCTACAGCAACGACCGTACCGTGGTTAGCGTTAAGACCGATAACAGGAAGGTG
TGAAAGCGGACGATACCCCGCGCAAAAGGAAACCGGTCCCGCGGTGGACGATAGCAAGGTTACCTACGACAC
CATTCAGAGCAAAGTGCTGAAGGCGGTTATCGATCAAGCGTTTCCGCGTGTGAAAGAGTATACCTGAACGTT
40 CACACCCTGCCGGGTACAGTTACAGCAATTTAACAAGTGTTTATTAAACACCGTATACCCCGGAAGTGA
CCTATAAGAAAATTAAACGAAACACCGCGGATACCTGATGAAGCTGCGTGACGATGCGCACCTGATCAACGC
GGAAATGACCGTGCGTCTGCAGGTGGTTGATAACCAACTGCACTTCGACGTGACCAAAAATTGTTAACCACAAC
CAGGTTACCCCGGTCAAAAGATTGACGATGAGCGTAACTGCTGAGCACCATCAGCTTTCTGGGCAACGCGC
TGTTTAGCGTGAGCAGCGATCAAGCGGTTGCGAAGTTTGTATGGTGGCACCATGAGCAACAACCCACGTTAG
45 CGGTGACGATCACATCGATGTGACCAACCCGATGAAAGACCTGGCGAAGGGTTACATGTATGGCTTTGTAGC
ACCGACAAGTGGCGGCGGTGTGTGGAGCAACAGCAGCAAAACAGCTACGGCGGTGGCAGCAACGATTGGACCC
GTCTGACCGCGTATAAAGAAACCGTTGGTAACGCGAATACGTGGGCATTACAGCAGCGAATGCGAGTGGGA
GAAAGCGTACAAGGGTATCGTGTTCGGAATATACCAAGGAGCTGCCGAGCGGAAAGTGGTTATACCCGAG
GATGCGAAGCGCGGACAACAAAGTGATTGGCAGGACCGTGCGATTGCGTACCGTAGCATCATGAACAACCCGC
50 AAGGCTGGGAAAAAGTTAAGGACATTACCGCGTATCGTATCGCGATGAACCTTTGGTAGCCAGGCGCAAAACCC
GTTCTGATGACCTGGACGGCATCAAGAAAATTAACTGCACACCGATGGCTGGGTGAGGCGGTTCTGCTG
AAGGGTTATGGTAGCGAGGGTCATGACAGCGGTACCTGAACCTACGCGGATATCGGTAACGATTTGGTGGCG
TGGAAGACTTTAAGACCTGATTGAGAAAGCGAAGAAATACGGTGGCGACCTGGGCATCCACGTTAACGCGAG
CGAAACCTACCCGGAGAGCAAGTATTTCAACGAAAACATTCTGCGTAAAAACCCGGACGGTAGCTACAGCTAT
55 GGTGGAACCTGGTGGATCAGGGTATCAACATTGATGCGGCGTACGACCTGGCGATGGCCGTCTGGCGCGTT
GGGAGGACCTGAAGAAAAAGCTGGGTGAAGGCTGGATTTTATCTATGTTGACGTGTGGGGTAACGGTCAGAG
CGGTGATAACGGTGGTGGCGACCATGTGCTGGCGAAAGAGATTAACAAGCAAGGTTGGCGTTTTGCGATC
GAATGGGGCCACGGTGGCGAGTACGACAGCACCTTCCGACCTGGGCGCGGATCTGACCTACGGTGGCTATA

CCAACAAGGGTATCAACAGCGGATTACCCGTTTCATCCGTAACCACCAGAAAGATAGCTGGGTTGGCGACTA
CCGTAGCTATGGTGGCGCGGCGAACTACCCGCTGCTGGGTGGCTATAGCATGAAGGACTTTGAGGGTTGGCAA
GGCCGTAGCGATTACAACGGTTATGTTACCAACCTGTTCGCGCACGACGTGATGACCAAGTACTTTCAGCACT
TCACCGTTAGCAAATGGGAAAACGGTACCCCGGTGACCATGACCGATAACGGCAGCACCTATAAGTGGACCCC
5 GGAAATGAAAGTGGAGCTGGTTGACGCGGCGGGTAACAAGGTGGTTGTGACCCGTAAAAGCAACGATGTGAAC
AGCCCGCAGTACCGTGAGCGTACCGTTACCCTGAACGGTCGTGTGATCCAAGACGGCAGCGCGTATCTGACCC
CGTGGAACCTGGGATGCGAACGGTAAAAAGCTGCCGACCGAAAAAGAGAAGATGTACTATTTTAAACACCCAAGC
GGGTGCGACCACTGGACCTGCCGAGCGACTGGGCGAACAGCAAGGTTTACCTGTATAAACTGACCGATCAG
GGCAAGACCGAGGAGCAAGAACTGACCGTGACCGATGGCAAAATTACCCTGGACCTGCTGGCGAACACGCGGT
10 ACGTTCTGTATCGTAGCAAGCAAACCAACCCGGAATGAGCTGGAGCGAGGGTATGCACATCTACGACCAAGG
TTTCAACAGCGCACCTGAAACACTGGACCATTAGCGGCGATGCGAGCAAGGCGGAGATCGTGAAAAGCCAG
GGTGCGAACGAAATGCTGCGTATCCAAGGCAACAAAAGCAAGGTTAGCCTGACCCAGAAGCTGACCGGTCTGA
AACCGAACACCAAGTACGCGGTTTATGTGGGCGTTGACAACCGTAGCAACGCGAAAGCGAGCATTACCGTTAA
CACCGGTGAAAAAGAGGTGACCACCTACACCAACAAGAGCCTGGCGCTGAACCTACATCAAAGCGTATGCGCAC
15 AACAACCGTCTGTGAGAACGCGACCGTGGACGATACCAAGCTACTTCCAGAACATGTATGCGTTCTTTACCACCG
GTAGCGACGTGAGCAACGTTACCCTGACCTGAGCCGTGAAGCGGGCGATGAGGCGACCTATTTTGACGAAAT
TCGTACCTTCGAGAACAACAGCAGCATGTACGGTGATAAGCACGACACCGGTCAGGGCACCTTTAAACAAGAT
TTCGAAAACGTTGCGCAAGGTATCTTCCCGTTTGTGTGGGTGGCGTGGAAGGCGTTGAGGACAACCGTACCC
ACCTGAGCGAGAAGCACGATCCGTACACCCAGCGTGGTTGGAACGGCAAAAAGGTGGACGATGTTATTGAGGG
20 TAACTGGAGCTGAAAACCAACGGCCTGGTTAGCCGTCGTAACTGGTGTACCAGACCATCCCGCAAAACTTC
CGTTTGTAGGCGGGCAAGACCTACCGTGTGACCTTTGAATATGAGGCGGGCAGCGACAACACCTATGCGTTTG
TTGTGGGTAAAGGCGAATTCAGAGCGGTCGTGTCGGCACCCAAGCGAGCAACCTGGAAATGCACGAGCTGCC
GAACACCTGGACCGATAGCAAAAAGGCGAAAAAGGTGACCTTCCTGGTTACCGGTGCGGAAACCGGTGACACC
TGGGTGGGTATCTACAGCACCGGCAACGCGAGCAACACCCGTGGTGATGCGGGTGGAACGCGAACTTTCTGTG
25 GCTATAACGATTTTCATGATGGACAACCTGCAAATCGAAGAGATTACCCTGACCGGCAAGATGCTGACCGAAAA
CGCGCTGAAAAACTATCTGCCGACCGTTGCGATGACCAACTACACCAAGGAAAGCATGGACGCGCTGAAAGAG
GCGGTTTTCAACCTGAGCCAGGCGGACGATGACATCAGCGTGGAAGAGGCGCGTGCGGAAATCGCGAAGATTG
AGGCGCTGAAAAACGCGCTGGTTTCAGAAAAAGACCGCGCTGGTTGCGGAAGATTTTGAGAGCCTGGATGCGCC
GGCGCAACCGGGTGAAGGCTGGAGAACCGGTTTCGACGGTAACGTTAGCAGCCTGTGGCACACCAAGCTGGAAC
30 GGTGGCGATGTTGGCAAGCCGGCGACCATGGTGCTGAAAGAACCGACCGAGATCACCGGTCTGCGTTATGTGC
CGCGTGCGAGCGATAGCAACGGCAACCTGCGTGACGTTAAGCTGGTTGTGACCGATGAAAGCGGTAAAGAGCA
CACCTTTAACGTGACCGACTGGCCGAACAACAACAAACCGAAGGATATTGACTTCGGCAAAACCATTAAGGCG
AAAAAGATCGTTCTGACCGGTACCAAGACCTACGGCGATGGTGGCGACAAATATCAGAGCGCGGCGGAGCTGA
TCTTTACCCGTCCGCAAGTGGCGGAAACCCCGCTGGATCTGAGCGGTTACGAAGCGGCGCTGGCGAAAGCGCA
35 GAAGCTGACCGATAAGGACAACCAGGAAGAGGTGGCGAGCGTTCAAGCGAGCATGAAATATGCGACCGACAAC
CACCTGCTGACCGAACGTATGGTTGCGTACTTTCGCGGATTATCTGAACCAACTGAAGGATAGCGCGACCAAAC
CGGATGCGCCGACCGAGCAGCAAGGGTGAAGAACAGCCGCCGCTGCTGGATGTTCCGGAGTTTTAAAGGTGGCGT
GAACGCGACCGAGGCGGCGGTGCACGAAGTTCGCGAGTTCAAGGGTGGCGTGAACGCGGTTACGCGCTGGTT
CACGAACTGCCGGAGTATAAAGGTGGCGCGAACGCGGTTCTGGCGGCGGCGAACGAAGTGCAGGAGTACAAGG
40 GTGGCGCGAACGCGGTGGAAGCGCTGGTTAACGAGAAACCGGCGTATACCGGTGTTCTGGCGACCGCGGGCGA
CCAGGCGGCGCGACCGTGGAAAAACCGGAGTACCCGCTGACCCGAGCCCGGTTGCGGACACCAAAACCCCG
GGTGCGAAAGATGAAGAGAAGCTGCCGCGGGTAGCGGCTCGAGCACCACCACCACCACCTGA

45

CLAIMS

1. A composition comprising a first sialidase which is independently selected from:
- (a) a polypeptide comprising or consisting of an amino acid sequence of SEQ ID NO: 2;
 - (b) a polypeptide comprising or consisting of an amino acid sequence which is at least
 - 5 85% identical to the amino acid sequence of SEQ ID NO: 2 or
 - (c) a polypeptide comprising or consisting of an amino acid sequence which is a fragment of the sequence of SEQ ID NO: 2 or a fragment of an amino acid which is 85% identical to the amino acid sequence of SEQ ID NO: 2;
- optionally wherein said first sialidase includes an additional methionine at the N terminus
- 10 and/or a histidine tag at the C terminus, which tag may be joined to the C terminus by a linker.
2. The composition according to claim 1 which further comprises a second sialidase which is independently selected from:
- 15 (d) a polypeptide comprising or consisting of an amino acid sequence of SEQ ID NO: 5;
 - (e) a polypeptide comprising or consisting of an amino acid sequence which is at least 85% identical to the amino acid sequence of SEQ ID NO: 5 or
 - (f) a polypeptide comprising or consisting of an amino acid sequence which is a fragment of the sequence of SEQ ID NO: 5 or a fragment of an amino acid which is
 - 20 85% identical to the amino acid sequence of SEQ ID NO: 5.
- optionally wherein said second sialidase includes an additional methionine at the N terminus and/or a histidine tag at the C terminus, which tag may be joined to the C terminus by a linker.
- 25 3. The composition according to claim 2, which is capable of hydrolysing >90% of sialic bonds in a glycoprotein, preferably a glycoprotein in its native, non-denatured state.
4. The composition according to any one of the preceding claims, which comprises the sialidase consisting of the amino acid sequence of SEQ ID NO: 3 and the sialidase
- 30 consisting of the amino acid sequence of SEQ ID NO: 6, preferably in a 1:1 ratio.

5. The composition according to any one of the preceding claims, wherein the first sialidase and/or the second sialidase are present in highly purified or isolated form; and/or wherein no more than two polypeptides in the composition are sialidases obtained from *Akkermansia muciniphila*.
- 5 6. The composition according to any one of the preceding claims, which additionally comprises a glycosidase and/or a protease, which is optionally present in highly purified or isolated form.
- 10 7. The composition according to claim 6, wherein said glycosidase is an O-glycosidase selected from:
- a. a polypeptide comprising or consisting of an amino acid sequence of SEQ ID NO: 9; or
 - b. a polypeptide comprising or consisting of an amino acid sequence which is
15 at least 85% identical to SEQ ID NO: 9;
 - c. a polypeptide comprising or consisting of an amino acid sequence which is a fragment of the sequence of SEQ ID NO: 9 or a fragment of a sequence which is 85% identical to the sequence of SEQ ID NO: 9; or
 - d. An O-glycosidase from *Enterococcus faecalis*, optionally a polypeptide
20 having the amino acid sequence of Uniprot entry B5UB72 version 22;
- optionally wherein said glycosidase consists of the amino acid sequence of SEQ ID NO: 10.
8. The composition according to claim 6 or 7, wherein said protease is an O-glycoprotein-specific endoprotease selected from:
- a. a polypeptide comprising or consisting of an amino acid sequence of SEQ ID NO: 11;
 - b. a polypeptide comprising or consisting of an amino acid sequence which is
25 at least 85% identical to the amino acid sequence of SEQ ID NO: 11 or
 - c. a polypeptide comprising or consisting of an amino acid sequence which is
30 a fragment of the sequence of SEQ ID NO: 11 or a fragment of a sequence which is 85% identical to the amino acid sequence of SEQ ID NO: 11;

optionally wherein said protease consists of the amino acid sequence of SEQ ID NO: 12.

9. A method for modifying a glycoprotein comprising contacting a sample containing the glycoprotein with a composition according to any one of the preceding claims,
5 optionally wherein the resulting products are analysed.

10. The method according to claim 9, wherein said analysis includes the separation and/or detection and/or isolation of the products by any suitable means, including SDS-PAGE, HPLC, lectin blotting, ELISA or mass spectrometry.

- 10
11. A method according to claim 9 or 10, wherein the composition is as defined in any one of claims 1 to 4, and wherein the method comprises contacting said sample with a protease or a glycosidase simultaneously, before or after the glycoprotein is contacted with said composition.

- 15
12. A method according to claim 9, wherein said glycosidase is an O-glycosidase, optionally as defined in claim 7; and/or said protease is an endoprotease, optionally as defined in claim 8.

20