

MODIFIED PROTEINS

SEQUENCE LISTING

5 The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on August 7, 2023, is named 70385US01P_SL.xml and is 124,012 bytes in size.

FIELD OF THE INVENTION

10 The present invention relates to the field of modified proteins, immunogenic compositions and vaccines comprising the modified proteins, their manufacture and the use of such compositions in medicine. More particularly, it relates to a modified Sap2 (Secreted Aspartyl Proteinase 2 of *Candida albicans*) protein. The modified Sap2 protein can be used as a carrier protein for other antigens, particularly saccharide antigens or other antigens lacking T cell epitopes.

BACKGROUND TO THE INVENTION

15 The impact of fungal infections on humans is a serious public health issue that has received much less attention than bacterial infection and treatment, despite ever-increasing incidence exacerbated by an increased incidence of immunocompromised individuals in the population. Fungal infections are also detrimental to the well-being of grazing livestock, with milk production in dairy
20 cows, and body and coat condition adversely affected by fungal infections. *Candida*, *Cryptococcus*, *Aspergillus*, and *Pneumocystis* are the most common fungal genera causing invasive human infections. *Candida* species, in particular, cause some of the most prevalent fungal infections and is the leading fungal pathogen worldwide. *Candida* species are early colonizers acquired at or near human birth primarily by physical contact. These organisms are able to colonize the skin, as well as the
25 gastrointestinal, and reproductive tracts of humans [9,10]. However, in certain situations (*e.g.* situations of immunosuppression), these organisms can become pathogenic and can cause a broad spectrum of human infections. During the last decades, the incidence of human infections caused by *Candida* genus has increased significantly (Sobel, 2007; Pf Fuller , 2011). *Candida* infections can be superficial or invasive. Invasive candidiasis are difficult to treat. Globally, an estimated 700,000
30 persons a year suffer from invasive candidiasis, with an associated mortality that may exceed 50%. Furthermore, *Candida* can also cause mucocutaneous infections, such as vulvovaginal candidiasis which, while rarely lethal, are associated with significant morbidity. Candidiasis have an estimated economical burden of \$3 billion in medical costs.

Candida albicans is the most studied member of the genus and is the most common
35 opportunistic pathogen and cause of invasive fungal infection in hospitalized human patients (Sobel, 2007; Pf Fuller , 2011). *C. albicans* is a highly adaptable fungal species that is prevalent in nosocomial

infections, and immunocompromised individuals are particularly at risk. *C. albicans* has a large repertoire of virulence factors that allows its transition from commensal organism (yeast form) to pathogen (hyphal form). For example, *C. albicans* is a commensal in the vaginal epithelium. Certain environmental conditions trigger the morphogenesis of *C. albicans* from the commensal yeast form to the pathogenic hyphal form. The hyphal form starts entering the lumen and breaches mucosal barriers, which leads to the symptomatic infection, e.g., vulvovaginal candidiasis (VVC) in women. Worldwide, approximately 70% - 80% of all women will be affected at least once in their lifetime by VVC, with *C. albicans* causing >85% of these infections. Up to 10% of these women will suffer from recurrence. Recurrent VVC or RVVC (defined as > 3 episodes/year) is more serious and refractory to therapy. RVVC results in reduced quality of life, a strong negative impact on work and social life, as well as increased associated healthcare costs. By 2030, the population of women with RVVC is estimated to increase to 158 million.

Despite the high mortality and morbidity rate of invasive fungal infections (IFIs), treatment options are limited and the anti-fungal drug discovery pipeline is under developed (Perfect, 2017). In addition, antifungal therapies are often ineffective. Moreover, the widespread use of antifungals in agriculture and clinical settings has led to the emergence of multidrug-resistant fungal infections (e.g. multidrug resistant *Candida auris* and other non-*albicans* species), which represent major threats to food security and human health (Nguyen et al., 2021). In addition, despite the substantial global burden of human fungal infections, there are no licensed vaccine currently available to prevent and/or control human fungal infections (Oliviera et al., 2021). Reasons for the lack of clinically available fungal vaccines include the very high costs related to manufacturing of the antigens and toxicology issues. In addition, fungal vaccines composed of whole organism (live-attenuated or killed fungal cells) show high immunogenicity, which makes them most likely to generate protective response but there is a potential risk of infection. Therefore, there is a clear and unmet need to identify novel targets and develop new classes of anti-fungal agents, including vaccines against pathogenic fungi (e.g., *C. albicans*) that can safely be produced in high quantities.

Given their theoretical safety advantages, identification of protective antigens for use in subunit vaccines has been the focus of much research. A major difference between the Fungi and Animalia kingdoms is the presence of cell wall on almost all fungal cells. The complex fungal cell wall is vital for maintaining cellular shape and integrity (Garcia-Rubio et al., 2020). An analysis of cell wall composition across a range of *Candida* species, including *C. albicans*, shows that the inner cell wall is composed mainly of β -glucans (β -1,3 and β -1,6 linked polymers of glucose), while the outer cell wall is comprised of highly glycosylated cell wall proteins that are decorated with N- and O-linked terminal mannans (branched polymer of mannose linked via α -1,2, α -1,3, α -1,4, α -1,6 and β -1,2 glycosidic bonds) (Ahmadipour et al., 2021, The Cell Surface, 7:100063). Consequently, these polysaccharide

components of the fungal cell wall, β -glucans and mannans, could be interesting candidates for developing safe and efficacious subunit fungal vaccines. However, polysaccharides are T-independent antigens that elicit antibody production via B lymphocytes without involvement of T-cells. Polysaccharides may elicit a long-lasting T-cell-dependent immune response in humans if they are coupled to a protein carrier that contains T-cell epitopes. Indeed, conjugation of T-independent antigens to carrier proteins has been established as a way of enabling T-cell help to become part of the immune response for a normally T-independent antigen. In this way, an immune response can be enhanced by allowing the development of immune memory and boost stability of the response. Successful conjugate vaccines against prokaryotic pathogenic bacteria have been developed by conjugating bacterial capsular saccharides to carrier proteins, the carrier protein having the known effect of turning the T-independent saccharide antigen into a T-dependent antigen capable of triggering an immune memory response. However, to date, it has not been possible to develop a successful conjugate vaccine against eukaryotic pathogenic fungi such as *Candida* by conjugating fungal cell wall polysaccharides to fungal carrier proteins. The present invention provides, for the first time, a new class of eukaryotic *Candida* glycoconjugate vaccines produced using the process of bioconjugation in a prokaryotic bacteria, namely *Escherichia coli*.

SUMMARY OF THE INVENTION

A glycoconjugate is a hybrid molecule composed of a carrier protein and multiple polysaccharide chains, wherein the polysaccharides are covalently linked to the carrier protein. This linkage of the antigenic polysaccharide to a carrier protein has brought significant advances in the field of vaccinology, eliciting a T-cell-dependent response characterized by the induction of immunological memory and improved immunogenicity. The standard approach to production of glycoconjugate vaccines is a chemical conjugation process that necessitates long development timelines, as the process requires extensive optimization for each individual target antigen. Additionally, the complexity of the production process results in high costs for such products. In contrast, bioconjugation is an innovative technology that allows the production of glycoconjugate vaccines in a biological environment (e.g., *E. coli*) to preserve native immunogenic structures. Using recombinant DNA technologies, the *E. coli* glycan biosynthesis machinery is genetically modified to produce the target polysaccharide antigen and covalently link it to an asparagine residue in a consensus sequence on a carrier protein. Thus, the glycoconjugate vaccine is produced entirely in *E. coli* in a single-step process, resulting in advantages for process reproducibility and robustness, while decreasing manufacturing cost. Using this bioconjugation technology, several glycoconjugate vaccines against prokaryotic organisms such as Gram negative bacteria (e.g. *Shigella*) and Gram positive bacteria (e.g. *Staphylococcus aureus* or *S. pneumoniae*) have been developed and tested in clinical trials. However, because *Candida* is a eukaryote, it has not been straightforward to transfer the biochemical pathway required for bioconjugation and production of *Candida* glycoproteins into *E.*

coli (as was previously done for bacterial lipopolysaccharides or capsular saccharides). Consequently, due to the higher complexity of the *Candida* bioconjugation process, a *Candida* glycoconjugate vaccine in *E. coli* has not previously been successfully produced.

5 The present invention provides modified Secreted Aspartyl Proteinase 2 (Sap2) proteins from *C. albicans* comprising at least one consensus sequence for glycosylation (e.g. D/E-X-N-Z-S/T) for use in conjugation to an antigen (e.g. *Candida* polysaccharide). The present invention further provides a glycoconjugate comprising the modified Sap2 carrier protein linked to the *Candida* polysaccharide antigen at one or more asparagine residues on the modified Sap2 protein, as well as methods of producing the *Candida* glycoconjugate vaccine in a host cell (e.g., *E. coli*). The present invention
10 further provides methods of preventing and/or treating RVVC using the *Candida* glycoconjugate vaccine.

Accordingly, there is provided in certain aspects of the present invention, a modified Secreted Aspartyl Proteinases 2 (Sap2) protein comprising amino acid residues 19-398 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to
15 amino acid residues 19-398 of SEQ ID NO: 1, modified in that the amino acid sequence comprises one or more consensus sequences comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.

According to further aspects of the invention, there is provided a modified Sap2 protein of the invention, wherein the modified Sap2 protein further comprises a substitution at amino acid residue
20 274 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1, optionally wherein the protein comprises an Aspartic Acid (D) to Asparagine (N) substitution at amino acid residue 274 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%,
25 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

According to further aspects of the invention, there is provided a modified Sap2 protein of the invention, wherein the modified Sap2 protein further comprises at least one Fructose biphosphate aldolase-1 (Fba) peptide comprising an amino acid sequence of YGKDVKDLFDYAQE (SEQ ID NO: 3) or an amino acid sequence at least 70%, 80%, 85%, 90%, or 92% identical to SEQ ID NO: 3.

30 According to further aspects of the invention, there is provided a modified Sap2 protein of the invention comprising an amino acid sequence of SEQ ID NO: 9.

According to further aspects of the invention, there is provided a modified Sap2 protein of the invention comprising an amino acid sequence of SEQ ID NO: 10.

According to further aspects of the invention, there is provided a conjugate (*e.g.* bioconjugate) comprising a modified Sap2 protein of the invention and at least one saccharide antigen.

According to further aspects of the invention, there is provided a modified Sap2 protein of protein of *Candida albicans* comprising (or consisting of): (1) an amino acid sequence of SEQ ID NO: 9 or SEQ ID NO: 10; and (2) at least one saccharide antigen of *Candida*, wherein the at least one saccharide antigen is a β -1,2 mannan polymer consisting of at least five consecutive β -1,2 linked mannose molecules, and wherein the at least one saccharide antigen is linked to at least one of four asparagine residues at positions 45, 94, 215, and 415 of SEQ ID NO: 9 or at least one of four asparagine residues at positions 6, 55, 176, and 376 of SEQ ID NO: 10.

According to further aspects of the invention, there is provided a polynucleotide encoding a modified Sap2 protein of the invention.

According to further aspects of the invention, there is provided a vector comprising a polynucleotide encoding a modified Sap2 protein of the invention.

According to further aspects of the invention, there is provided a host cell comprising: (1) one or more polynucleotide sequences that encode one or more heterologous glycosyltransferases; (2) a polynucleotide sequence that encodes a heterologous oligosaccharyl transferase; (3) a polynucleotide sequence that encodes a modified Sap2 protein of the invention; and, optionally, (4) a polynucleotide sequence that encodes a polymerase.

According to further aspects of the invention, there is provided a method for producing a bioconjugate that comprises (or consists of) a modified Sap2 protein linked to at least one saccharide antigen, the method comprising: (1) culturing a host cell of the invention under conditions suitable for the production of proteins; and (2) isolating the bioconjugate produced by said host cell, optionally isolating the bioconjugate from a periplasmic extract from the host cell.

According to further aspects of the invention, there is provided an immunogenic composition comprising a modified Sap2 protein of the invention, a conjugate of the invention, or a bioconjugate of the invention and optionally a pharmaceutically acceptable excipient and/or carrier.

According to further aspects of the invention, there is provided a method of making the immunogenic composition of the invention, the method comprising the step of mixing a modified Sap2 protein of the invention, a conjugate of the invention, or a bioconjugate of the invention, with a pharmaceutically acceptable excipient or carrier.

According to further aspects of the invention, there is provided a vaccine comprising an immunogenic composition of the invention and, optionally, a pharmaceutically acceptable excipient or carrier and optionally an adjuvant.

According to further aspects of the invention, there is provided a *Candida albicans* vaccine comprising: (1) a modified Sap2 protein of the invention; (2) at least one *Candida albicans* saccharide antigen linked to said modified Sap2 protein; and, optionally, (3) a pharmaceutically acceptable carrier or adjuvant.

5 According to a further aspect of the invention, there is provided a method for treatment or prevention of *Candida albicans* infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention.

10 According to further aspects of the invention, there is provided a method for immunizing a subject (*e.g.* human) against *Candida albicans* infection, the method comprising administering to the subject an immunoprotective dose of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention.

15 According to further aspects of the invention, there is provided a method of inducing immune response to *Candida albicans* infection in a subject (*e.g.* human), the method comprising administering to the subject a therapeutically or prophylactically effective amount of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention.

20 According to further aspects of the invention, there is provided a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention, for use in treatment or prevention of a disease caused by *Candida albicans* infection.

25 According to further aspects of the invention, there is provided a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention for use in manufacture of a medicament for treatment or prevention of a disease caused by *Candida albicans* infection.

According to further aspects of the invention, there is provided a host cell comprising:

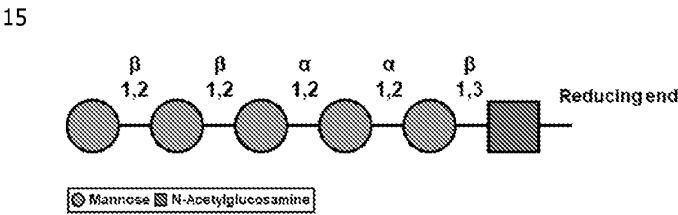
- 30 i. a nucleotide sequence encoding one or more first heterologous glycosyltransferase(s) capable of synthesizing a β -1,2 mannan polymer;
- ii. A nucleotide sequence encoding a second heterologous glycosyltransferase which is eukaryotic and is capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain;
- iii. a nucleotide sequence encoding a heterologous oligosaccharyl transferase; and

- iv. optionally, a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.

According to further aspects of the invention, there is provided a method of producing a glycoconjugate comprising a modified carrier protein and a β -1,2 mannan, wherein said method comprises culturing the host cell of the invention under conditions suitable for the production of proteins.

According to further aspects of the invention, there is provided a method of producing a glycoconjugate comprising a modified carrier protein and a β -1,2 mannan, wherein said method comprises the steps of i) culturing a host cell of the invention under conditions suitable for the production of proteins, ii) harvesting the culture to produce a harvested culture, and iii) isolating the glycoconjugate from the culture.

According to further aspects of the invention, there is provided a saccharide that is a β -1,2 mannan polymer comprising the structure:



According to further aspects of the invention, there is provided a saccharide comprising the structure:



According to further aspects of the invention, there is provided a conjugate (*e.g.* bioconjugate) comprising a saccharide of the invention linked to an asparagine residue of a modified carrier protein.

DESCRIPTION OF DRAWINGS/FIGURES

FIGS. 1A and 1B show the structure of mature Sap2 protein from *C. albicans* of SEQ ID NO: 88 (comprising residues 57-398 of SEQ ID NO: 1). **FIG. 1A:** spheres indicate the positions for insertion of glycosites. **FIG. 1B:** spheres indicate the best positions for insertion of glycosites. N: glycosite inserted next to position 19; C: glycosite inserted next to position 398; Mut3a: glycosite substituted for positions 98-102; Mut 4d: glycosite substituted for positions 109-113; and Mut 8a: glycosite inserted next to position 220 (all positions are relative to wild-type full-length Sap2 sequence of *C. albicans* having SEQ ID NO: 1).

FIG. 2 shows production of a proSap2-Fba- β -1,2-mannan bioconjugate in *E. coli*. **FIG. 2A** shows the a schematic of the structure of the *C. Albicans* mannan. **FIG. 2B** shows the biosynthesis scheme for the modified Sap2-mannan bioconjugate in *E. coli*. Enterobacterial O-antigen cluster refers to *Citrobacter freundii* P079F I.

5 **FIG. 3** shows glycosylation tests with a series of modified proSap2 proteins, each comprising a single glycosite. **FIGS. 3A** and **3B** show SDS-PAGE analysis of modified proSap2 proteins purified from PPE by IMAC. Many of the modified proSap2 proteins showed very good glycosylation with Kp05. Modified proSap2 "C" variant (FIG. 3A, lane 3) shows the highest expression level and glycosylation. Some positions (e.g., Mut14a-14e (FIG. 3B, lanes27-31)) seem to destabilize the
10 modified proSap2 protein and lead to reduced protein expression.

FIG. 4 shows glycosylation tests with a series of modified proSap2 proteins comprising combined glycosites. Lanes 1-4 show SDS-PAGE analysis of modified proSap2 proteins (not comprising Fba peptide sequence) purified from PPE by IMAC. Lanes 5-8 show SDS-PAGE analysis of modified proSap2 proteins (comprising the Fba peptide sequence) purified from PPE by IMAC. As shown,
15 modified proSap2 proteins with up to 4 glycosites were expressed and purified.

FIG. 5 shows SDS-PAGE gels followed by coomassie blue staining (FIG. 5A) or Western Blot analyses (FIGS. 5B-5E) of purified modified Sap2 proteins. Lane M: protein standard; Lane 1: purified unglycosylated modified proSap2 (MutN3C-8a-Fba-C). Lane 2: purified modified proSap2-Fba- β -1,2-mannan bioconjugate (glycosylated MutN3C-8a-Fba-C). **FIG. 5A:** SDS-PAGE analysis
20 followed by coomassie blue staining. **FIG. 5B:** anti-proSap2 immunoblot (against N-terminal "pro-peptide" region); FIG. 5C: anti-Sap2 immunoblot (against peptide in the middle of the protein sequence). FIG. 5D: anti-Fba immunoblot; FIG. 5E: anti-mannan immunoblot.

FIG. 6 shows structural determination via crystallography modified proSap2 protein produced in *E. coli* superimposed onto published wild-type mature Sap2 (mSap2) protein of *C. albicans*. Spheres
25 represent boundaries of the residue range taken from the AlphaFold model.

FIG. 7 shows Circular dichroism (CD) spectroscopy characterization of wild-type proSap2 protein and modified proSap2- β -1,2-mannan bioconjugate. **FIG. 7A** shows near-UV CD spectroscopy analysis. **FIG. 7B** shows far-UV CD spectroscopy analysis. **Fig. 7C** shows the secondary structure element content after analysis with "CDNN" software (Applied Photophysics Ltd) for deconvolution of CD
30 spectroscopy.

FIG. 8 shows preclinical testing of a purified modified proSap2-Fba- β -1,2-mannan bioconjugate ("proSap2-4FM") in rabbit. **FIG. 8A** shows the modified proSap2-Fba- β -1,2-mannan bioconjugate attributes. **FIG. 8B** shows a 3D representation of the modified proSap2 protein-mannan bioconjugate; **FIG. 8C** shows the rabbit immunization scheme with the modified proSap2-Fba - β -
35 1,2-mannan bioconjugate.

FIG. 9 shows immunogenicity of the purified modified proSap2-Fba- β -1,2-mannan bioconjugate ("proSap2-4FM") in rabbits. **FIGS. 9A, 9B and 9C** show immunogenicity against Sap2, Fba and β -mannan, respectively, of the modified proSap2-Fba- β -1,2-mannan bioconjugate ("proSap2-4FM") compared to modified mature Sap2 protein ("mSap2") or modified Sap2-Fba- β -1,2-mannan bioconjugate ("Sap2-4FM^{unmodif}") in rabbit.

FIG. 10 shows a protease activity inhibition assay of full-length wild-type Sap2 protein of *C. albicans* tested with a strong inhibitor of its activity (a mono-specific Fab fragment). **FIG. 10A:** the measured activities in absence of Sap2 (negative control), in absence of the Fab fragment (positive control), and in presence of the Fab fragment (Fab anti-Sap2) were expressed as percentage of the positive control activity as determined via absorbance at 280 nm. The employed Fab fragment was able to completely block the protease activity of Sap2. **FIG. 10B:** titration curves are shown where sera from a bicomponent immunization ("Candi5V", where proSap2-Fba- β -1,2-mannan bioconjugate and Als3-Fba- β -1,3-glucan bioconjugate are formulated together) or proSap2-Fba- β -1,2-mannan rabbits immunization "proSap2-4FM") are used. As a control, sera from rabbits with mock immunization was used.

FIG. 11 shows the capacity of antibodies against proSap2-4FM bioconjugate to inhibit adhesion of *C. albicans* to vaginal epithelial cells.

FIG. 12 shows the capacity of antibodies against proSap2-4FM bioconjugate to mediate neutrophil killing of *C. albicans* hyphae.

FIG. 13 shows the capacity of antibodies against proSap2-4FM bioconjugate to bind to *C. albicans* using fluorescent microscopy.

DETAILED DESCRIPTION OF THE INVENTION

DEFINITIONS

As used herein, the term "Sap2 protein" or "Sap2" refers to a wild type Secreted Aspartyl Proteinase 2 protein comprising a wild type leader sequence (amino acid residues 1-18) and a pro-peptide sequence (amino acid residues 19-56) at its N-terminus. In certain embodiments, the Sap2 protein is from *Candida*, optionally from *Candida albicans*. In specific embodiments, the Sap2 protein comprises the amino acid sequence of SEQ ID NO.: 1.

As used herein, the term "proSap2 protein" or "proSap2" refers to a C-terminal fragment of a wild type Sap2 protein. In certain embodiments, the proSap2 protein is from *Candida*, optionally from *Candida albicans*. In specific embodiments, the proSap2 protein comprises amino acid residues 19-398 of SEQ ID NO.: 1. In some embodiments, the proSap2 protein comprises an amino acid sequence of SEQ ID NO.: 17.

As used herein, the term "modified" protein refers to a protein that is altered (in one or more way) as compared to wild type protein (e.g. a "modified Sap2 protein" or "modified Sap2" excludes a

wild type Sap2 protein and a "modified proSap2 protein" or "modified proSap2" excludes a wild type proSap2 protein). The term "modified Sap2 protein" refers to a Sap2 protein that comprises one or more consensus glycosite sequences of the invention (*e.g.* a consensus sequence comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline).embodiments, This, in certain embodiments, a modified Sap2 protein refers to a Sap2 protein that comprises one or more consensus glycosite sequences (*e.g.* comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline), wherein the one or more consensus sequences have been added next to or substituted for one or more amino acid residues of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. In other embodiments, a modified Sap2 protein refers to a Sap2 protein that comprises one or more consensus glycosite sequences (*e.g.* comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline), wherein the one or more consensus sequences have been added next to or substituted for one or more amino acid residues within amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1. In yet other embodiments, a modified Sap2 protein refers to a Sap2 protein that comprises one or more consensus glycosite sequences (*e.g.* comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline), wherein the one or more consensus sequences have been added next to or substituted for one or more amino acid residues within amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1. In specific embodiments, a modified Sap2 protein refers to an Sap2 protein that comprises amino acid residues 19-398 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1, modified in that the amino acid sequence comprises one or more consensus sequences (*e.g.* comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline), wherein the one or more consensus sequences have been added next to or substituted for one or more amino acid residues within amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1. In additional embodiments, a modified Sap2 protein further comprises a substitution at amino acid residue 88 and/or 274 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. In other embodiments, a modified Sap2 protein further comprises a substitution at amino acid residue 88 and/or 274 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at

least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1. In still other embodiments, a modified Sap2 protein further comprises a substitution at amino acid residue 88 and/or 274 of amino acid residues 57-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1). One of skill in the art will know that these residues correspond to of amino acid residue 32 and amino acid residue 218, respectively, of SEQ ID NO: 88 (i.e., a *Candida albicans* Sap2 protein lacking the wild type leader sequence and lacking the wild type propeptide sequence). In specific aspects, the substitution is a D274N substitution. In certain aspects, the substitution eliminates the biological activity of the modified Sap2 protein. In certain embodiments, a modified Sap2 protein of the invention is a naturally occurring modified Sap2 protein. In other embodiments, a modified Sap2 protein of the invention is a recombinant modified Sap2 protein. In yet other embodiments, a modified Sap2 protein of the invention is an isolated recombinant modified Sap2 protein. In certain embodiments, the modified Sap2 protein is from *Candida*, optionally from *Candida albicans*. In some embodiments, the modified Sap2 protein comprises an amino acid sequence of SEQ ID NO: 9. In other embodiments, the modified Sap2 protein comprises an amino acid sequence of SEQ ID NO: 10. In yet embodiments, the modified Sap2 protein comprises an amino acid sequence selected from, but not limited to, the amino acid sequences of N, N3C, C, Mut1a, Mut 1b, Mut2a, Mut 2b, Mut3a, Mut 3b, Mut4a, Mut4b, Mut4c, Mut4d, Mut4e, Mut5a, Mut5b, Mut6a, Mut6b, Mut7, Mut8a, Mut8b, Mut9, Mut10, Mut11, Mut12, Mut13, Mut14a, Mut14b, Mut14c, Mut14d, Mut14e, Mut15, Mut16, Mut17, Mut 18, Mut 19, and Mut20.

As used herein, the term "control Sap2 protein" means, without limitation,: (1) a Sap2 protein that does not comprise one or more consensus sequences inserted next to or substituted for one or more amino acids of SEQ ID NO: 1 (or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1) or added to the N-terminal and/or C-terminal end of SEQ ID NO: 1 (or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1); (2) a Sap2 protein that does not comprise one or more consensus sequences inserted next to or substituted for one or more amino acids of a Sap2 protein that comprises amino acid residues 19-398 of SEQ ID NO: 1 (or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1) or added to the N-terminal and/or C-terminal end of a Sap2 protein that comprises amino acid residues 19-398 SEQ ID NO: 1 (or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1); or (3) a Sap2 protein that does not comprise one or more consensus sequences inserted next to or substituted for one or more amino acids of a Sap2 protein that comprises amino acid residues 57-398 of SEQ ID NO: 1 (or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1) or added to the N-terminal and/or C-terminal end of a

Sap2 protein that comprises amino acid residues 57-398 SEQ ID NO: 1 (or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1);. Thus, a control Sap2 protein includes, without limitation, a wild type Sap2 protein, a wild type Sap2 protein of SEQ ID NO: 1, a wild type Sap2 protein comprising amino acid residues 19-398 of SEQ ID NO: 1, or a wild type Sap2 protein comprising amino acid residues 57-398 of SEQ ID NO: 1

As used herein, the term "carrier protein" refers to a protein which may be linked to an antigen (*e.g.* saccharide antigen, such as a fungal polysaccharide antigen) to create a conjugate (*e.g.* bioconjugate). A carrier protein activates T-cell mediated immunity in relation to the antigen to which it is conjugated. As used herein, the term "carrier protein" refers to a protein that comprises one or more consensus sequences to which a saccharide antigen of the invention is linked. In certain aspects, a carrier protein is a modified Sap2 protein of the invention.

As used herein, the term "any amino acid except proline (pro, P)" refers to an amino acid selected from the group consisting of alanine (ala, A), arginine (arg, R), asparagine (asn, N), aspartic acid (asp,D), cysteine (cys, C), glutamine (gln, Q), glutamic acid (glu, E), glycine (gly, G), histidine (his, H), isoleucine (ile,I), leucine (leu, L), lysine (lys, K), methionine (met, M), phenylalanine (phe, F), serine (ser, S), threonine (thr, T), tryptophan (trp, W), tyrosine (tyr, Y), and valine (val, V).

As used herein, the term "naturally occurring amino acid residues" refers to amino acids that are naturally incorporated into polypeptides. In particular, the 20 amino acids encoded by the universal genetic code: alanine (ala, A), arginine (arg, R), asparagine (asn, N), aspartic acid (asp,D), cysteine (cys, C), glutamine (gln, Q), glutamic acid (glu, E), glycine (gly, G), histidine (his, H), isoleucine (ile,I), leucine (leu, L), lysine (lys, K), methionine (met, M), phenylalanine (phe, F), proline (pro, P), serine (ser, S), threonine (thr, T), tryptophan (trp, W), tyrosine (tyr, Y), and valine (val, V).

As used herein, the term "glycosyltransferases (GTFs, Gtfs)" refers to enzymes that establish glycosidic linkages. Glycosyltransferases are enzymes that catalyze the formation of the glycosidic linkage to form a glycoside. For example, they catalyze the transfer of saccharide moieties from an activated nucleotide sugar (also known as the "glycosyl donor") to a nucleophilic glycosyl acceptor molecule, the nucleophile of which can be oxygen-, carbon-, nitrogen-, or sulfur-based.

As used herein, the term "oligosaccharyl transferases (OTases or OSTs)" refers to enzymes that catalyze a mechanistically unique and selective transfer of an oligo- or polysaccharide (glycosylation) to the asparagine (N) residue at the consensus sequence of nascent or folded proteins. OSTs transfer of a 14-sugar oligosaccharide from dolichol to nascent protein. OST is a type of glycosyltransferase. The reaction catalyzed by OST is the central step in the N-linked glycosylation pathway. OST is a component of the translocon in the endoplasmic reticulum (ER) membrane. A lipid-linked core-oligosaccharide is assembled at the membrane of the endoplasmic reticulum and

transferred to selected asparagine residues of nascent polypeptide chains by the oligosaccharyl transferase complex.

As used herein, the term "O-Antigens (also known as O-specific polysaccharides or O-side chains)" refers to a component of the surface lipopolysaccharide (LPS) of Gram-negative bacteria.

5 Examples include O-antigens from *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*.

As used herein, the term "Lipopolysaccharide (LPS)" refers to large molecules comprising a lipid and a polysaccharide joined by a covalent bond.

As used herein, the term "capsular polysaccharide (CP)" refers to a polysaccharide found on the bacterial cell wall. Examples include capsular polysaccharide from *Streptococcus pneumoniae*,
10 *Haemophilus influenzae*, *Neisseria meningitidis* and *Staphylococcus aureus*.

As used herein, the term "reducing end" refers to the reducing end of an oligosaccharide or polysaccharide is the monosaccharide with a free anomeric carbon that is not involved in a glycosidic bond and is thus capable of converting to the open-chain form.

15 As used herein, the term "conjugate" refers to a protein (e.g. a carrier protein) covalently linked to an antigen.

As used herein, the term "bioconjugate" refers to a conjugate between a protein (e.g. a carrier protein) and an antigen (e.g. a saccharide antigen, such as a bacterial polysaccharide antigen) prepared in a host cell background, wherein host cell machinery links the antigen to the protein (e.g. N-linked glycosylation). Usually, in a bioconjugate the polysaccharide is linked to asparagine via N-acetylglucosamine.
20

As used herein, the term "immunogenic fragment" refers to a portion of an antigen smaller than the whole, that is capable of eliciting a humoral and/or cellular immune response in a host animal, e.g. human, specific for that fragment. Fragments of a protein can be produced using techniques known in the art, e.g. recombinantly, by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Typically, fragments comprise at least 10, 20, 30, 40 or 50 contiguous amino acids of the full length sequence. Fragments may be readily modified by adding or removing 1, 2, 3, 4, 5,
25 6, 7, 8, 9, 10, 20, 30, 40 or 50 amino acids from either or both of the N and C termini. In certain aspects, a fragment of a modified Sap2 protein of the invention still comprises the recited modifications that are made to the Sap2 protein.

As used herein, the term "conservative amino acid substitution" involves substitution of a native amino acid residue with a non-native residue such that there is little or no effect on the size,
35 polarity, charge, hydrophobicity, or hydrophilicity of the amino acid residue at that position, and without resulting in decreased immunogenicity. For example, these may be substitutions within the

following groups: valine, glycine; glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Conservative amino acid modifications to the sequence of a polypeptide (and the corresponding modifications to the encoding nucleotides) may produce polypeptides having functional and chemical characteristics similar to those of a parental polypeptide.

As used herein, the term "deletion" refers to the removal of one or more amino acid residues from the protein sequence. Typically, no more than about from 1 to 6 residues (*e.g.* 1 to 4 residues) are deleted at any one site within the protein molecule.

As used herein, the terms "insertion" or "addition" (including other tenses thereof such as "inserted") refers to the addition of one or more non-native amino acid residues in the protein sequence or, as the context requires, addition of one or more non-native nucleotides in the polynucleotide sequence. Typically, no more than about from 1 to 10 residues, (*e.g.* 1 to 7 residues, 1 to 6 residues, or 1 to 4 residues) are inserted at any one site within the protein molecule.

As used herein, the term "added next to" refers to the addition of one or more non-native amino acid residues in the protein sequence at a position adjacent to the referenced amino acid or amino acid region. For example, "added next to one or more amino acids between amino acid residues 19-24" means the addition at a position adjacent to any one of amino acid residues 19-24 (including adjacent to amino acid residues 19 or 24).

As used herein, the term "glycosite" refers to an amino acid sequence recognized by a bacterial oligosaccharyl transferase, *e.g.* PglB of *Campylobacter jejuni*. Thus, a glycosite refers to an amino acid sequence within a carrier protein (*e.g.* a modified Sap2 protein of the invention or a modified proSap2 protein of the invention) to which an antigen saccharide (*e.g.* a β -1,2 mannan polymer) is linked, either covalently or non-covalently.

A "consensus sequence" refers to a sequence have a specific structure and/or function. As used herein, the term "consensus sequence" is a sequence comprising a glycosite. A consensus sequence of the invention includes, but is not limited to, a five amino acid consensus sequence D/E-X-N-Z-S/T, a seven amino acid consensus sequence K-D/E-X-N-Z-S/T-K, and an extended consensus sequence (*e.g.* J-U-B-D/E-X-N-Z-S/T-J-U-B).

As used herein, the term "introduced at" is used herein to reference the location and manner of inserting a consensus sequence into an amino acid sequence. A consensus sequence (or glycosite) which is introduced at an N-terminal or C-terminal position of a protein may be added next to the amino acid sequence at the N-terminus or C-terminus, whereas a consensus sequence (or glycosite) which is introduced at a specific amino acid residue within the protein (*e.g.* amino acid residue 19 of SEQ ID NO: 1), may be substituted for that amino acid.

Unless specifically stated otherwise, providing a numeric range (*e.g.* "19-24") is inclusive of endpoints (*i.e.* includes the values 19 and 24). For example, "between amino acids 19 to 398 of SEQ

ID NO: 1" refers to position in the amino acid sequence between amino acid 19 and amino acid 398 of SEQ ID NO: 1 including both amino acids 19 and 398.

The term "identical" or percent "identity" refers to nucleotide sequences or amino acid sequences that are the same or have a specified percentage of nucleotide residues or amino acid residues that are the same (*e.g.* 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identity over a specified region), when compared and aligned for maximum correspondence using, for example, sequence comparison algorithms or by manual alignment and visual inspection. Identity between polypeptides may be calculated by various algorithms. In general, when calculating percentage identity the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include inserting "gaps" in either one or both sequences, to enhance the degree of alignment. For example the Needleman Wunsch algorithm (Needleman and Wunsch 1970, J. Mol. Biol. 48: 443-453) for global alignment, or the Smith Waterman algorithm (Smith and Waterman 1981, J. Mol. Biol. 147: 195- 197) for local alignment may be used, *e.g.* using the default parameters (Smith Waterman uses BLOSUM 62 scoring matrix with a Gap opening penalty of 10 and a Gap extension penalty of 1). A preferred algorithm is described by Dufresne *et al.* in Nature Biotechnology in 2002 (vol. 20, pp. 1269-71) and is used in the software GenePAST (Genome Quest Life Sciences, Inc. Boston, MA). The GenePAST "percent identity" algorithm finds the best fit between the query sequence and the subject sequence, and expresses the alignment as an exact percentage. GenePAST makes no alignment scoring adjustments based on considerations of biological relevance between query and subject sequences. Identity between two sequences is calculated across the entire length of both sequences and is expressed as a percentage of the reference sequence (*e.g.* SEQ ID NO: 1 of the invention).

As used herein the term "recombinant" means artificial or synthetic. In certain embodiments, a "recombinant protein" refers to a protein that has been made using recombinant nucleotide sequences (nucleotide sequences introduced into a host cell). In certain embodiments, the nucleotide sequence that encodes a "recombinant protein" is heterologous to the host cell.

As used herein the term "isolated" or "purified" refers to a protein, conjugate (*e.g.* bioconjugate), polynucleotide, or vector in a form not found in nature. This includes, for example, a protein, conjugate (*e.g.* bioconjugate), polynucleotide, or vector having been separated from host cell or organism (including crude extracts) or otherwise removed from its natural environment. In certain embodiments, an isolated or purified protein is a protein essentially free from all other polypeptides with which the protein is innately associated (or innately in contact with).

As used herein, the term "subject" refers to an animal, in particular a mammal such as a primate (*e.g.* human).

As used herein, the term "therapeutically or prophylactically effective amount," in the context of administering a therapy (*e.g.* an immunogenic composition or a vaccine of the invention) to a

subject refers to the amount of a therapy which has a prophylactic and/or therapeutic effect(s). In certain embodiments, an "therapeutically or prophylactically effective amount" refers to the amount of a therapy which is sufficient to achieve one, two, three, four, or more of the following effects: (i) reduce or ameliorate the severity of a fungal infection or symptom associated therewith; (ii) reduce the duration of a fungal infection or symptom associated therewith; (iii) prevent the progression of a fungal infection or symptom associated therewith; (iv) cause regression of a fungal infection or symptom associated therewith; (v) prevent the development or onset of a fungal infection, or symptom associated therewith; (vi) prevent the recurrence of a fungal infection or symptom associated therewith; (vii) reduce organ failure associated with a fungal infection; (viii) reduce hospitalization of a subject having a fungal infection; (ix) reduce hospitalization length of a subject having a fungal infection; (x) increase the survival of a subject with a fungal infection; (xi) eliminate a fungal infection in a subject; (xii) inhibit or reduce a fungal replication in a subject; and/or (xiii) enhance or improve the prophylactic or therapeutic effect(s) of another therapy.

As used herein, the term "immunoprotective dose," in the context of administering a therapy (e.g. an immunogenic composition or a vaccine of the invention) to a subject refers to the amount of a therapy which has a prophylactic and/or therapeutic effect(s). In certain embodiments, an "immunoprotective dose" refers to the amount of a therapy which is sufficient to achieve one, two, three, four, or more of the following effects: (i) reduce or ameliorate the severity of a fungal infection or symptom associated therewith; (ii) reduce the duration of a fungal infection or symptom associated therewith; (iii) prevent the progression of a fungal infection or symptom associated therewith; (iv) cause regression of a fungal infection or symptom associated therewith; (v) prevent the development or onset of a fungal infection, or symptom associated therewith; (vi) prevent the recurrence of a fungal infection or symptom associated therewith; (vii) reduce organ failure associated with a fungal infection; (viii) reduce hospitalization of a subject having a fungal infection; (ix) reduce hospitalization length of a subject having a fungal infection; (x) increase the survival of a subject with a fungal infection; (xi) eliminate a fungal infection in a subject; and/or (xii) inhibit or reduce a fungal replication in a subject.

The term "comprises" is open-ended and means "includes." Thus, unless the context requires otherwise, the word "comprises" or "has", and variations thereof (including "comprise" and "comprising" or "have" and "having", respectively), will be understood to imply the inclusion of a stated compound(s), molecule(s), composition(s), or steps, but not to the exclusion of any other compound(s), molecule(s), composition(s), or steps. The terms "comprising" and "having" when used as a transition phrase herein are open-ended whereas the term "consisting of" when used as a transition phrase herein is closed (*i.e.*, limited to that which is listed and nothing more). In certain embodiments and for readability, the word "is" may be used as a substitute for "consists of" or "consisting of". The abbreviation, "*e.g.*" is derived from the Latin *exempli gratia*, and is used herein

to indicate a non-limiting example. Thus, the abbreviation “*e.g.*” is synonymous with the term “for example”.

The term “conjugate vaccine” refers to a vaccine created by covalently linking a polysaccharide antigen to a carrier protein. Conjugate vaccine elicit immune response against a pathogen (*e.g.* a
5 fungus) and immunological memory. In infants and elderly people, a protective immune response against polysaccharide antigens can be induced if these antigens are conjugated with proteins that induce a T-cell dependent response.

The term “glycoconjugate vaccine” refers to a vaccine comprising a protein carrier linked to an antigenic or immunogenic oligosaccharide.

10 As used herein, “undecaprenyl” or “und” refers to undecaprenol lipid composed of eleven prenol units. “Und-P” refers to undecaprenyl phosphate, which is a universal lipid carrier (derived from Und) of glycan biosynthetic intermediates for carbohydrate polymers. “Und-PP” refers to undecaprenyl pyrophosphate, which is a phosphorylated version of Und-P.

15 “Periplasmic space” or “periplasm” refers to the space between the inner cytoplasmic membrane and external outer membrane of a host cell (*e.g.* Gram-negative bacteria, *e.g.* *E. coli*).

The terms “of” and “from” are used herein interchangeably. Thus, a saccharide antigen “from” *Candida* means, without limitation, (1) a saccharide antigen obtained from *Candida* or (2) a saccharide antigen of *Candida*, i.e., a saccharide antigen comprising a structure similar to a saccharide antigen from *Candida*, but, *e.g.*, produced recombinantly in a host cell (*e.g.* a bacterial cell).

20 wbaB refers to a glycosyltransferase. In certain embodiments, wbaB is a glycosyltransferase obtained from an organism including, but not limited to, *Citrobacter freundii* P079F I, *Salmonella* O6,7 (C₁) Thompson, or *Escherichia coli* O17. In certain embodiments, wbaB is a glycosyltransferase of *Citrobacter freundii* P079F I. In some aspects, wbaB is a wild type glycosyltransferase. In other aspects, wbaB is a non-naturally occurring (*e.g.*, mutant and/or recombinant) glycosyltransferase.

25 wbaC refers to a glycosyltransferase. In certain embodiments, wbaC is a glycosyltransferase obtained from an organism including, but not limited to, *Citrobacter freundii* P079F I, *Salmonella* O6,7 (C₁) Thompson, or *Escherichia coli* O17. In certain embodiments, wbaC is a glycosyltransferase of *Citrobacter freundii* P079F I. In some aspects, wbaC is a wild type glycosyltransferase. In other aspects, wbaC is a non-naturally occurring (*e.g.*, mutant and/or recombinant) glycosyltransferase.

30 wbaD refers to a glycosyltransferase. In certain embodiments, wbaD is a glycosyltransferase obtained from an organism including, but not limited to, *Citrobacter freundii* P079F I, *Salmonella* O6,7 (C₁) Thompson, or *Escherichia coli* O17. In certain embodiments, wbaD is a glycosyltransferase of *Citrobacter freundii* P079F I. In some aspects, wbaD is a wild type glycosyltransferase. In other aspects, wbaD is a non-naturally occurring (*e.g.*, mutant and/or recombinant) glycosyltransferase.

Bmt3 refers to a glycosyltransferase. In certain embodiments, Bmt3 is a glycosyltransferase obtained from an organism including, but not limited to, *Candida albicans*. In certain embodiments, Bmt3 is a glycosyltransferase of *C. albicans*. In some aspects, Bmt3 is a wild type glycosyltransferase gene. In other aspects, Bmt3 is a non-naturally occurring (e.g., mutant and/or recombinant) glycosyltransferase gene. In specific embodiments, the Bmt3 gene is a codon-optimized version of the gene from *C. albicans*. In additional embodiments, the Bmt3 protein is N-terminally fused with the leader peptide sequence from *E. coli* OmpC for its periplasmic export.

manB refers to a guanylyltransferase. In certain embodiments, manB is a guanylyltransferase obtained from an organism including, but not limited to *Escherichia coli*. In certain embodiments, manB is a guanylyltransferase of *E. coli* K12 W3110. In some aspects, manB is a wild type guanylyltransferase. In other aspects, manB is a non-naturally occurring (e.g., mutant and/or recombinant) guanylyltransferase.

manC refers to a guanylyltransferase. In certain embodiments, manC is a guanylyltransferase obtained from an organism including, but not limited to *Escherichia coli*. In certain embodiments, manC is a guanylyltransferase of *E. coli* K12 W3110. In some aspects, manC is a wild type guanylyltransferase. In other aspects, manC is a non-naturally occurring (e.g., mutant and/or recombinant) guanylyltransferase.

PglB refers to a oligosaccharyl transferase. In certain embodiments, pglB is a oligosaccharyl transferase obtained from an organism including, but not limited to, *Campylobacter jejuni*, *Campylobacter coli*, or *Sinorhizobium meliloti* 1021. In certain embodiments, pglB is a oligosaccharyl transferase of *Campylobacter coli*. In some aspects, pglB is a wild type oligosaccharyl transferase. In other aspects, PglB is a non-naturally occurring oligosaccharyl transferase. In some aspects, the pglB gene is an evolved oligosaccharyl transferase gene. In additional aspects, the pglB protein is an evolved pglB, i.e., an evolved oligosaccharyl transferase. By "evolved" is meant a protein or nucleic acid that has undergone directed evolution. Directed evolution is a method used in protein engineering that mimics the process of natural selection to steer proteins or nucleic acids toward a user-defined goal. The process of directed evolution consists of subjecting a gene to iterative rounds of mutagenesis (creating a library of variants), selection (expressing those variants and isolating members with the desired function) and amplification (generating a template for the next round). It can be performed in vivo (in living organisms), or in vitro (in cells or free in solution). Directed evolution is used both for protein engineering as an alternative to rationally designing modified proteins, as well as for experimental evolution studies of fundamental evolutionary principles in a controlled, laboratory environment. In certain embodiments, the evolved pglB of the invention comprises an amino acid sequence of SEQ ID NO: 16. Thus, in certain embodiments, the pglB amino acid sequence in SEQ ID NO:16 contains one or more mutations which enhances the activity of pglB for the saccharide antigen of the invention. Thus, in certain aspects, the evolved pglB of the invention transfers a saccharide

antigen of the invention to a modified Sap2 protein of the invention more efficiently in comparison to a wild-type pglB (e.g., a wild type pglB obtained from *Campylobacter jejuni*).

Wzx refers to a translocase. In certain embodiments, wxz is a translocase obtained from an organism including, but not limited to, *Citrobacter freundii* P079F I, *Salmonella* O6,7 (C₁) Thompson, or *Escherichia coli* O17. In certain embodiments, wxz is a translocase of *Citrobacter freundii* P079F I. In some aspects, wxz is a wild type translocase. In other aspects, wxz is a non-naturally occurring (e.g., mutant and/or recombinant) translocase.

Sap2 Protein

10 Secreted Aspartyl Proteinase 2 protein of *Candida albicans* (also known as "Sap2") is an extracellular virulence factor that is the predominant enzyme in vaginal secretion of *Candida* infected symptomatic women. Sap2 provides nutrition for the *Candida* cells, facilitates attachment to host tissue, facilitates fungal epithelial and endothelial penetration, contributes to *Candida's* capability to evade immune responses, and is immunogenic during infection (Kumar R. et al., 2015, Infect Immun, 15 83(7):2614-2626). Sap2 functions as a hydrolytic enzyme, exhibiting broad substrate specificity, and is expressed abundantly in a culture of *C. albicans* (Naglik JR, et al, 2003, Microbiol. Mol. Biol. Rev., 67:400-428). The Secreted Aspartyl Proteinase (Sap) proteins of are encoded by a family of 10 *SAP* genes and have been the most comprehensively studied as key virulence determinants of *C. albicans* (Id.). Sap2 is a member of the secreted aspartyl Sap family of proteins and is encoded by the *SAP2* 20 gene. All 10 *SAP* genes of *C. albicans* encode preproenzymes approximately 60 amino acids longer than the mature enzymes, which are processed when transported via the secretory pathway (Id.). The mature enzymes contain sequence motifs typical for all aspartyl proteinases, including the two conserved aspartate residues of the active site and conserved cysteine residues implicated in the maintenance of the three-dimensional structure (Id.).

25 The *SAP2* gene encodes a 398-amino acid long preproprotein that is processed to a 342-residue mature enzyme, a typical aspartic proteinase of pH optimum 3-4, displaying sensitivity to pepstatin A, a peptide-based inhibitor (Cutfield S., et al., 1995, Structure, 3:1261-1271). At the N terminus of the Sap2 preproprotein is a 18-amino acid long signal peptide ("SP" or "leader sequence") followed by a 38-amino acid long pro-peptide.

30 During the maturation process, the *C. albicans* Sap2 preproprotein is processed to a 380-residue "intermediate" form ("proSap2"; SEQ ID NO: 17) that lacks the leader sequence, and then further processed to the 342-residue "mature" form ("mSap2"; SEQ ID NO: 88) that lacks the leader sequence and the pro-peptide sequence (Naglik et al. 2003). Within the C-terminal domain there is a structurally distinct subdomain of about 100 residues (Figs. 1A and 1B), that in some other aspartic 35 proteinases has been shown to behave as a separate rigid body. This domain in *SAP2* comprises residues 197–215 and 223–305, and contains one disulphide bond (between 256 and 294) which ties

together a double loop (243–255, 282–293) of random structure. These C-terminal loops are a peripheral feature of aspartic proteinases, forming part of the wide entrance to the binding site. Another loop, formed by the disulphide 47–59, is more well defined. It also flanks the binding site but is considerably closer. The pattern of secondary structure elements in SAP2, in particular the organisation of the many β strands and the two major α -helical sections (140–146, 228–237), is similar to other known aspartic proteinases. However, some of the turns linking these elements are different in conformation. A highly conserved feature of the aspartic proteinases is the “flap” region (a β hairpin loop) which interacts centrally with bound inhibitor/substrate, shielding the active site from bulk solvent. In SAP2 the flap comprises residues Lys81–Gln91, with the tyrosine at position 84 being highly conserved, as found in other aspartic proteinases (equivalent to Tyr75 in pepsin) (Cutfield S., et al., 1995, *Structure*, 3:1261-1271). The overall structure of Sap2 conforms to the classical aspartic proteinase fold typified by pepsin. One of the most noticeable properties of Sap2 is the variety of proteins it can cleave. Sap2 is known to degrade many human proteins including molecules that protect mucosal surfaces such as mucin and secretory immunoglobulin A (IgA). Sap2 can also degrade molecules of the extracellular matrix such as keratin, collagen and vimentin (Naglik JR, et al, 2003, *Microbiol. Mol. Biol. Rev.*, 67:400-428). A Sap2 protein useful in the invention can be produced by methods known in the art in view of the present disclosure, see for example Smolenski G, Sullivan PA, Cutfield SM, Cutfield JF., 1997, *Microbiology*, 143 (Pt 2):349-356.

As shown below, a full-length wild-type Sap2 protein of *C. Albicans* comprises the amino acid sequence of SEQ ID NO: 1 (with wild-type leader sequence underlined and the pro-peptide italicized; the amino acid residue (D) at position 274 is double underlined; in certain aspects, substitution of this residue (e.g. to N) eliminates the biological activity of Sap2):

MFLKNIFIALAIALLVDA *TPTTTKRSAGFVALDFS*VVKTPKAFPVTNGQEGKTSKRQAVPVLHNEQVTYAADITV
GSNNQKLNIVIDTGSDDLWVPDVNVDCQVTYSDQTADFCQKQGTYPGSSASQDLNTPFKIGYGDGSSSQG
TLYKDTVGGVSIKNQVLADVDSIDQILGVGYKTNEAGGSYDNPVTLKKQGVIAKNAYSLYLNSPDAATG
QIIFFGGVDNAKYSGSLIALPVTSDRELRLISLGSVEVSGKTINTDNVDVLVDSGTTITYLQQDLADQIIKAFNGKLT
QDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGDNFLRSAYIV
YDLDDNEISLAQVKYTSASSISALT (SEQ ID NO: 1)

In certain embodiments, the present invention provides a modified Sap2 protein. The term “modified Sap2 protein” refers to a Sap2 protein comprising an amino acid sequence (for example, comprising an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, or comprising an amino acid sequence of amino acid residues 19-398 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1), which Sap2 amino acid sequence has been modified by the

addition, substitution or deletion of one or more amino acids (for example, by addition of a consensus sequence(s) selected from D/E-X-N-Z-S/T, K-D/E-X-N-Z-S/T and/ or an extended consensus sequence (e.g. J-U-B-D/E-X-N-Z-S/T-J-U-B); and/or by substitution of one or more amino acids by a consensus sequence(s) selected from D/E-X-N-Z-S/T, K-D/E-X-N-Z-S/T. For example, a modified Sap2 protein

5 may be a Sap2 protein having an amino acid sequence of SEQ ID NO: 1 which has been modified in that the amino acid sequence comprises one or more consensus sequences selected from from D/E-X-N-Z-S/T, K-D/E-X-N-Z-S/T and/ or an extended consensus sequence (e.g. J-U-B-D/E-X-N-Z-S/T-J-U-B). As used herein, in consensus sequences of the present invention, X and Z are independently any amino acid except proline; preferably, X is Q (glutamine) and Z is A (alanine). In certain embodiments,

10 a modified Sap2 protein of the invention may comprise further modifications (e.g., additions, substitutions, and/or deletions of one or more amino acid residues). In specific embodiments, a modified Sap2 protein of the invention comprises amino acid residues 19-398 of SEQ ID NO.: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1, modified in that the amino acid sequence comprises

15 one or more consensus sequences comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline. In other aspects, a modified Sap2 protein of the invention comprises amino acid residues 57-398 of SEQ ID NO.: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1, modified in that the amino acid sequence comprises one or more consensus

20 sequences comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.

In some embodiments, the modified Sap2 protein of the invention is a non-naturally occurring Sap2 protein (*i.e.* not native). In certain embodiments, a modified Sap2 protein of the invention may have an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical

25 to SEQ ID NO: 1. In some aspects, a modified Sap2 protein of the invention may have an amino acid sequence at least 80% identical to SEQ ID NO: 1. In other aspects, a modified Sap2 protein of the invention may have an amino acid sequence at least 85% identical to SEQ ID NO: 1. In yet other aspects, a modified Sap2 protein of the invention may have an amino acid sequence at least 90% identical to SEQ ID NO: 1. In still other aspects, a modified Sap2 protein of the invention may have an amino acid sequence at least 91% identical to SEQ ID NO: 1. In additional aspects, a modified

30 Sap2 protein of the invention may have an amino acid sequence at least 92% identical to SEQ ID NO: 1. In other aspects, a modified Sap2 protein of the invention may have an amino acid sequence at least 93% identical to SEQ ID NO: 1. In still other aspects, a modified Sap2 protein of the invention may have an amino acid sequence at least 94% identical to SEQ ID NO: 1. In yet other aspects, a

35 modified Sap2 protein of the invention may have an amino acid sequence at least 95% identical to SEQ ID NO: 1. In certain aspects, a modified Sap2 protein of the invention may have an amino acid

sequence at least 96% identical to SEQ ID NO: 1. In other aspects, a modified Sap2 protein of the invention may have an amino acid sequence at least 97% identical to SEQ ID NO: 1. In additional aspects, a modified Sap2 protein of the invention may have an amino acid sequence at least 98% identical to SEQ ID NO: 1. In yet other aspects, a modified Sap2 protein of the invention may have
5 an amino acid sequence at least 99% identical to SEQ ID NO: 1.

In certain aspects, a modified Sap2 protein of the invention comprises one or more consensus glycosite sequences. The terms "glycosite sequence", "consensus glycosite sequence" and "consensus sequence" are used herein interchangeably. In certain aspects, a modified Sap2 protein of the invention comprises at least one, at least two, at least three, at least four, at least five, at least six,
10 at least seven, at least eight, at least nine, or at least ten consensus sequences. In other aspects, a modified Sap2 protein of the invention contains one, two, three, four, five, six, seven, eight, nine, or ten consensus sequences. In specific aspects, a modified Sap2 protein of the invention contains at least three consensus sequences. In preferred aspects, a modified Sap2 protein of the invention contains three consensus sequences. In certain embodiments, a modified Sap2 protein of the
15 invention comprises one or more consensus sequences, wherein all of the consensus sequences have identical amino acid sequences. In other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein all of the consensus sequences have different amino acid sequences. In yet other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein at least two of the consensus sequences have identical
20 amino acid sequences.

Thus, in certain embodiments, the present invention provides a modified Sap2 protein comprising amino acid residues 19-398 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1, modified in that the amino acid sequence comprises one or more consensus sequences
25 comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.

In certain embodiments, the modified Sap2 protein of the invention is an inactive protein. By "inactive" protein is meant that the protein lacks any biological activity. In certain aspects, the modified Sap2 protein of the invention comprises a substitution at one or more positions selected from
30 the group consisting of amino acid residue 88 of SEQ ID NO: 1 and amino acid residue 274 of SEQ ID NO: 1. One of skill in the art will know that these residues correspond to of amino acid residue 32 and amino acid residue 218, respectively, of SEQ ID NO: 88 (i.e., a *Candida albicans* Sap2 protein lacking the wild type leader sequence and lacking the wild type propeptide sequence). In certain aspects, a substitution of one or both amino acid residues at positions 88 and 274 of SEQ ID NO: 1
35 renders the Sap2 protein inactive. In specific aspects, a substitution of residue 274 of SEQ ID NO: 1

renders the Sap2 protein inactive (e.g., eliminates the hydrolytic activity of Sap2). In certain aspects, elimination of the hydrolytic activity of Sap2 eliminates the virulence of the protein.

Thus, in certain embodiments, a modified Sap2 protein of the invention further comprises a substitution at amino acid residue 274 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1. In certain aspects, a modified Sap2 protein of the invention comprises an Aspartic Acid (D) to Asparagine (N) substitution at amino acid residue 274 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1. In certain aspects, the substitution renders the modified Sap2 protein inactive.

In certain aspects, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have each been added next to or substituted for one or more amino acids, selected from specific amino acid residues within a modified Sap2 protein of the invention (consensus sequence sites). These one or more consensus sequence sites are independently selected from (1) one or more amino acids between amino acid residues 19-24 (*e.g.* amino acid residue 19), (2) one or more amino acids between amino acid residues 52-62 (*e.g.* amino acid residue 57), (3) one or more amino acids between amino acid residues 93-107 (*e.g.* one or more amino acids between amino acid residues 98-102), (4) one or more amino acids between amino acid residues 104-118 (*e.g.* one or more amino acids between amino acid residues 109-113), (5) one or more amino acids between amino acid residues 142-144 (*e.g.* amino acid residue 143), (6) one or more amino acids between amino acid residues 187-197 (*e.g.* amino acid residue 192), (7) one or more amino acids between amino acid residues 215-225 (*e.g.* amino acid residue 220), (8) one or more amino acids between amino acid residues 243-253 (*e.g.* amino acid residue 248), (9) one or more amino acids between amino acid residues 255-265 (*e.g.* amino acid residue 260), (10) one or more amino acids between amino acid residues 263-273 (*e.g.* amino acid residue 268), (11) one or more amino acids between amino acid residues 320-330 (*e.g.* amino acid residue 325), (12) one or more amino acids between amino acid residues 339-349 (*e.g.* amino acid residue 344), (13) one or more amino acids between amino acid residues 358-359 (*e.g.* amino acid residue 358 or 359), and (14) one or more amino acids between amino acid residues 388-398 (*e.g.* amino acid residue 398) of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

In other aspects, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for one or more amino acids selected from the group consisting of: one or more amino acids between amino acid residues 19-24 (*e.g.* amino acid residue 19), one or more amino acids between amino acid residues 52-62 (*e.g.* amino acid residue 57), one or more amino acids between amino acid residues

93-107 (*e.g.* one or more amino acids between amino acid residues 98-102), one or more amino acids between amino acid residues 104-118 (*e.g.* one or more amino acids between amino acid residues 109-113), one or more amino acids between amino acid residues 142-144 (*e.g.* amino acid residue 143), one or more amino acids between amino acid residues 187-197 (*e.g.* amino acid residue 192),
5 one or more amino acids between amino acid residues 215-225 (*e.g.* amino acid residue 220), one or more amino acids between amino acid residues 243-253 (*e.g.* amino acid residue 248), one or more amino acids between amino acid residues 255-265 (*e.g.* amino acid residue 260), one or more amino acids between amino acid residues 263-273 (*e.g.* amino acid residue 268), one or more amino acids between amino acid residues 320-330 (*e.g.* amino acid residue 325), one or more amino acids
10 between amino acid residues 339-349 (*e.g.* amino acid residue 344), one or more amino acids between amino acid residues 358-359 (*e.g.* amino acid residue 358 or 359), and (14) one or more amino acids between amino acid residues 388-398 (*e.g.* amino acid residue 398) of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of
15 SEQ ID NO: 1.

In additional aspects, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for one or more amino acids selected from the group consisting of: one or more amino acids between amino acid residues 52-62 (*e.g.* amino acid residue 57), one or more amino acids
20 between amino acid residues 93-107 (*e.g.* one or more amino acids between amino acid residues 98-102), one or more amino acids between amino acid residues 104-118 (*e.g.* one or more amino acids between amino acid residues 109-113), one or more amino acids between amino acid residues 142-144 (*e.g.* amino acid residue 143), one or more amino acids between amino acid residues 187-197 (*e.g.* amino acid residue 192), one or more amino acids between amino acid residues 215-225 (*e.g.*
25 amino acid residue 220), one or more amino acids between amino acid residues 243-253 (*e.g.* amino acid residue 248), one or more amino acids between amino acid residues 255-265 (*e.g.* amino acid residue 260), one or more amino acids between amino acid residues 263-273 (*e.g.* amino acid residue 268), one or more amino acids between amino acid residues 320-330 (*e.g.* amino acid residue 325), one or more amino acids between amino acid residues 339-349 (*e.g.* amino acid residue 344), one or
30 more amino acids between amino acid residues 358-359 (*e.g.* amino acid residue 358 or 359), and (14) one or more amino acids between amino acid residues 388-398 (*e.g.* amino acid residue 398) of amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1.

35 In certain embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or

substituted for amino acid residue 19 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

5 In other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 57 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

10 In some embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for one or more amino acids between amino acid residues 109-113 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

15 In yet other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 220 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

20 In certain embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 19 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

25 In other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 57 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

30 In some embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for one or more amino acids between amino acid residues 109-113 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

35 In yet other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 220 of amino acid residues 19-398 of SEQ ID NO: 1 or at an

equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

In still other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 398 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

In additional embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 57 of amino acid residues 57-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1.

In some embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for one or more amino acids between amino acid residues 109-113 of amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1.

In yet other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 220 of amino acid residues 57-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1.

In still other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 398 of amino acid residues 57-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1.

In specific embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 57; and (ii) amino acid residue 398 of amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1. In certain embodiments, a modified Sap2 protein of the invention comprises the amino acid sequence of SEQ ID NO: 81. In other embodiments, a modified Sap2 protein of the invention comprises the amino acid sequence of SEQ ID NO: 85 (which additionally includes the fba sequence).

In specific embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 57; (ii) amino acid residue 220; and (iii) amino acid residue 398 of amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1. In certain embodiments, a modified Sap2 protein of the invention comprises the amino acid sequence of SEQ ID NO: 82. In other embodiments, a modified Sap2 protein of the invention comprises the amino acid sequence of SEQ ID NO: 86 (which additionally includes the fba sequence).

10 In specific embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 57; (ii) the amino acids between amino acid residues 98-102; (iii) amino acid residue 220; and (iv) amino acid residue 398 of amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 15 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1. In certain embodiments, a modified Sap2 protein of the invention comprises the amino acid sequence of SEQ ID NO: 83. In other embodiments, a modified Sap2 protein of the invention comprises the amino acid sequence of SEQ ID NO: 87 (which additionally includes the fba sequence). In certain embodiments, a modified Sap2 protein of the invention comprising the amino acid sequence of SEQ ID NO: 87 was 20 selected as the optimal protein carrier because it led to all the introduced glycosites being at least partially glycosylated, resulting in only minimal unglycosylated protein left and maximizing the sugar/protein ratio (see Fig. 4; MutN3C-3a-8a-fba-C).

Thus in certain embodiments, the present invention provides a method of increasing efficiency of glycosylation of a modified Sap2 protein of the invention, the method comprising expressing in a 25 host cell of the invention a modified Sap2 protein comprising at least four consensus sequences, wherein the consensus sequences have been added next to or substituted for (i) amino acid residue 57; (ii) the amino acids between amino acid residues 98-102; (iii) amino acid residue 220; and (iv) amino acid residue 398 of amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical 30 to amino acid residues 57-398 of SEQ ID NO: 1, wherein glycosylation efficiency of the modified Sap2 protein is increased relative to glycosylation efficiency of a control Sap2 protein not comprising the at least four consensus sequences. In some embodiments, the modified Sap2 protein comprises the amino acid sequence of SEQ ID NO: 83. In other embodiments, the modified Sap2 protein comprises the amino acid sequence of SEQ ID NO: 87.

35 In other embodiments, the present invention provides a method of increasing efficiency of glycosylation of a modified Sap2 protein of the invention, the method comprising expressing in a host

cell of the invention a modified Sap2 protein comprising at least four consensus sequences, wherein the consensus sequences have been added next to or substituted for (i) amino acid residue 19; (ii) the amino acids between amino acid residues 98-102; (iii) amino acid residue 220; and (iv) amino acid residue 398 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1, wherein glycosylation efficiency of the modified Sap2 protein is increased relative to glycosylation efficiency of a control Sap2 protein not comprising the at least four consensus sequences.

In further embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 220 and (ii) amino acid residue 398 of amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1. In certain embodiments, a modified Sap2 protein of the invention comprises the amino acid sequence of SEQ ID NO: 84.

In yet other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 57; (ii) the amino acids between amino acid residues 98-102; and (iii) amino acid residue 220 of amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1.

In further embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 57; (ii) the amino acids between amino acid residues 98-102; (iii) the amino acids between amino acid residues 109-113; (iv) amino acid residue 220; and (v) amino acid residue 398 of amino acid residues 57-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 57-398 of SEQ ID NO: 1.

In certain embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 19; and (ii) amino acid residue 398 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

In other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 19; (ii) amino acid residue 220; and (iii) amino acid residue

398 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

5 In yet other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 19; (ii) the amino acids between amino acid residues 98-102; (iii) amino acid residue 220; and (iv) amino acid residue 398 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

10 In still other embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 220 and (ii) amino acid residue 398 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

15 In further embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 19; (ii) the amino acids between amino acid residues 98-102; and (iii) amino acid residue 220 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

20 In additional embodiments, a modified Sap2 protein of the invention comprises one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for (i) amino acid residue 19; (ii) the amino acids between amino acid residues 98-102; (iii) the amino acids between amino acid residues 109-113; (iv) amino acid residue 220; and (v) amino acid residue 398 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

25 In certain embodiments, the modified Sap2 protein of the invention is from a fungi. In certain aspects, the fungi is *Candida*. Thus, in some embodiments, the modified Sap2 protein of the invention is from *Candida*. In certain aspects, the *Candida* includes, but is not limited to, *Candida albicans*, *Candida auris*, *Candida guilliermondii*, *Candida lusitanaea* and *Candida tropicalis*. In specific aspects, the modified Sap2 protein of the invention is from *Candida albicans*. In certain embodiments, at least one of the one or more consensus sequences comprises an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline. In certain aspects, X is Q (glutamine). In other aspects, Z is A (alanine). In some embodiments, the one or more consensus sequences include, but are not limited to, DQNT (SEQ ID NO: 4) and DQNV (SEQ ID NO: 5). In

specific embodiments, X is Q (glutamine), Z is A (alanine) and the one or more consensus sequences are selected from the group consisting of DQNAT (SEQ ID NO: 4) and DQNVT (SEQ ID NO: 5).

In certain embodiments, a modified Sap2 protein of the invention further comprises at least one Fructose biphosphate aldolase (Fba) peptide. The Fba peptide is a 14-mer peptide derived from the N-terminal portion of fructose-bisphosphate aldolase protein (Fba-1). Fba-1 protein is a multifunctional *C. albicans* cell wall protein and an important enzyme of glycolytic pathway. It can facilitate fungal attachment to human cells or abiotic surfaces, and protects *Candida* cells from the host's immune system (Elamin E, et al., 2021, *J. Immunol. Res.*, 2021:1-19). In addition to this, it promotes the detoxification of reactive oxygen species generated during respiratory burst. Proteomics analysis revealed that Fba1 is the most abundant and stable enzyme in *Candida* and is considered to be one of the main immunodominant proteins (Elamin E, et al., 2021, *J. Immunol. Res.*, 2021:1-19). The Fba peptide has been previously been used to generate a self-adjuvanting vaccine (Xin H et al., 2012, *PLoS ONE*, 7:e35106). In certain aspects, the at least one Fba peptide comprises (or consists of) an amino acid sequence of YGKDVKDLFDYAE (SEQ ID NO: 3). In other aspects, the at least one Fba peptide comprises (or consists of) an amino acid sequence at least 70%, 80%, 85%, 90%, or 92% identical to SEQ ID NO: 3.

In specific embodiments, the modified Sap2 protein of the invention comprises at least one Fba peptide. In certain embodiments, the at least one Fba peptide comprises an amino acid sequence of YGKDVKDLFDYAE (SEQ ID NO: 3) or an amino acid sequence at least 70%, 80%, 85%, 90%, or 92% identical to SEQ ID NO: 3. In certain embodiments, the at least one Fba peptide is linked to a modified Sap2 protein of the invention. In some embodiments, the at least one Fba peptide is non-covalently linked to a modified Sap2 protein of the invention. In other embodiments, the at least one Fba peptide is covalently linked to a modified Sap2 protein of the invention. In additional embodiments, the Fba peptide is linked to a modified Sap2 protein of the invention at a single amino acid residue. In other embodiments, the Fba peptide is linked to a modified Sap2 protein of the invention at more than one amino acid residues. In additional embodiments, a Fba peptide is linked to a modified Sap2 protein of the invention at one or more amino acid residues. In specific aspects, the Fba peptide is linked to a modified Sap2 protein of the invention at amino acid residue 398. The numbering of the amino acid residues as specified herein refers to the amino acid positions in SEQ ID NO: 1 (or where an amino acid sequence is at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 equivalent positions to that of SEQ ID NO: 1 if this sequence was lined up with an amino acid sequence of SEQ ID NO: 1 in order to maximise the sequence identity between the two sequences). In certain embodiments, the Fba peptide is covalently linked to a modified Sap2 protein of the invention at amino acid residue 398 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%,

modified Sap2 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

In certain embodiments, a modified Sap2 protein of the invention comprises at least one additional consensus sequence. In some embodiments, the at least one additional consensus sequence has been added next to C-terminal amino acid residue of a modified Sap2 protein of the invention. In other embodiments, a modified Sap2 protein of the invention comprises at least one Fba peptide and the at least one additional consensus sequence has been added next to C-terminal amino acid residue of the at least one Fba peptide. In additional embodiments, a modified Sap2 protein of the invention comprises an Fba peptide comprising an amino acid sequence of YGKDVKDLFDYAE (SEQ ID NO: 3) or an amino acid sequence at least 70%, 80%, 85%, 90%, or 92% identical to SEQ ID NO: 3, and the at least one additional consensus sequence has been added next to C-terminal amino acid residue of SEQ ID NO: 3 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 3. In specific embodiments, the modified Sap2 protein comprises at least one additional consensus sequence comprising an amino acid sequence of J-U-B-D/E-X-N-Z-S/T-J-U-B, wherein X and Z are independently any amino acid except proline and J, U and B independently comprise 1 to 5 naturally occurring amino acid residues, wherein the at least one additional consensus sequence has been added next to C-terminal amino acid residue of SEQ ID NO: 3 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 3. In certain embodiments, the at least one additional consensus sequence comprises an amino acid sequence of J-U-B-D/E-X-N-Z-S/T-J-U-B, wherein X and Z are independently any amino acid except proline and J, U and B independently comprise one or more naturally occurring amino acid residues. In certain aspects, X is Q (glutamine). In other aspects, Z is A (alanine). In certain embodiments, J comprises at least one Glycine (G) residue. In some aspects, J comprises 1 to 10, 1 to 9, 1 to 8, 1 to 7, 1 to 6, 1 to 5, 1 to 4, 1 to 3, or 1 to 2 Glycine (G) residues. In specific aspects, J comprises 1 to 5 Glycine (G) residues. In other aspects, B comprises at least one Glycine (G) residue. In certain aspects, B comprises 1 to 10, 1 to 9, 1 to 8, 1 to 7, 1 to 6, 1 to 5, 1 to 4, 1 to 3, or 1 to 2 Glycine (G) residues. In specific aspects, B comprises 1 to 5 Glycine (G) residues. In additional embodiments, each of J and B comprises 1 to 5 Glycine (G) residues. In additional aspects, U comprises at least one Glycine (G) residue. In certain aspects, U comprises 1 to 10, 1 to 9, 1 to 8, 1 to 7, 1 to 6, 1 to 5, 1 to 4, 1 to 3, or 1 to 2 Glycine (G) residues. In specific aspects, U comprises 1 to 5 serine (S) residues. In specific embodiments, X is Q, Z is A, each of J and B comprises 1 to 5 Glycine (G) residues and U comprises 1 to 5 serine (S) residues. In specific embodiments, the additional consensus sequence comprises (or consists of) an amino acid sequence of GSGGGDQNATGSGGG (SEQ ID NO: 7). In other embodiments, the additional consensus sequence

comprises (or consists of) an amino acid sequence of GSGGGDQ~~Q~~NATGSGGGHHHHHHHHHH (SEQ ID NO: 8).

In certain embodiments, a modified Sap2 protein of the invention comprises (or consists of) an amino acid sequence of SEQ ID NO: 9:

5 STPTTTKRSAGFVALDFSVVKTPKAFPVTNGOEGKTSKR **GGG**~~DQ~~NAT *GGG*QAVPVT~~LH~~NEQVTYAAD
ITVGSNNQKLNIVDTGSSDLWV**P**~~DQ~~**NVT**CQVTYSDQTADFCQKGTYPGSSASQDLNTPFKIGYGDGSSS
QGTLYKDTVFGGGVSIKNQVLADVSTSIDQGILGVGYKTNEAGGSYDNPVTLKKQGVIAKNAYS~~LY~~LN**SP**~~DQ~~
NATGQIIFGGVDNAKYSGSLIALPVTSDRELRLISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQIIKAFN
GKLTQDSNGNSFYEVDCNLSGDVWFNFSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGDNFLRS
10 AYIVYDLDDNEISLAQVKYTSASSISALTYGKDVKDLFDYAQEGSGG**GDQ**NATGSGGG

SEQ ID NO: 9 represents an intermediate form of modified Sap2 protein, lacking the native leader sequence, comprising the pro-peptide sequence (undelined), comprising 4 glycosites (bolded), comprising an inactivating substitution (double underlined), and further comprising an Fba sequence (dashed undelined).

15 In other embodiments, a modified Sap2 protein of the invention comprises (or consists of) an amino acid sequence of SEQ ID NO: 10:

GGG~~DQ~~NAT *GGG*QAVPVT~~LH~~NEQVTYAADITVGSNNQKLNIVDTGSSDLWV**P**~~DQ~~**NVT**CQVTYSDQ
TADFCQKGTYPGSSASQDLNTPFKIGYGDGSSSQGTLYKDTVFGGGVSIKNQVLADVSTSIDQGILGVGY
KTNEAGGSYDNPVTLKKQGVIAKNAYS~~LY~~LN**SP**~~DQ~~NATGQIIFGGVDNAKYSGSLIALPVTSDRELRLISLGSVE
20 VSGKTINTDNVDVLVNSGTTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVWFNFSKNAKISVPAS
EFAASLQGDDGQPYDKCQLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTYGKDVKDLFD
YAQEGSGG**GDQ**NATGSGGG

SEQ ID NO: 10 represents a mature form of modified Sap2 sequence, lacking the native leader sequence or the pro-peptide sequence, comprising 4 glycosites (bolded), comprising an inactivating
25 substitution (double underlined), and further comprising an Fba sequence (dashed undelined).

In certain embodiments, a modified Sap2 protein of the invention is glycosylated. In certain aspects, the modified Sap2 protein of the invention is N-glycosylated.

It will be understood by a person skilled in the art, that reference to "between amino acids ..." (for example "between amino acids 19-24") is referring to the amino acid number counting
30 consecutively from the N-terminus of the amino acid sequence, for example "between amino acids 19 to 24...of SEQ ID NO: 1" refers to position in the amino acid sequence between amino acid 19 and amino acid 24 of SEQ ID NO: 1 including both amino acids 19 and 24. Thus, in certain embodiments, where "one or more consensus sequences have been added next to or substituted for one or more amino acids between amino acid residues 19-24 of SEQ ID NO: 1", the one or more consensus

sequences may have been added next to or substituted for any one (or more) of amino acid numbers 19, 20, 21, 22, 23, and 24 in SEQ ID NO: 1. A person skilled in the art will understand that when the Sap2 amino acid sequence is a variant and/or fragment of an amino acid sequence of SEQ ID NO: 1, such as an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, the reference to "between amino acids ..." refers to the position that would be equivalent to the defined position, if this sequence was lined up with an amino acid sequence of SEQ ID NO: 1 in order to maximise the sequence identity between the two sequences (Sequence alignment tools are not limited to Clustal Omega ([www\(.\).ebi\(.\).ac\(.\).uk](http://www.ebi.ac.uk)) MUSCLE ([www\(.\).ebi\(.\).ac\(.\).uk](http://www.ebi.ac.uk)), or T-coffee ([www\(.\).tcoffee\(.\).org](http://www.tcoffee.org)). In one aspect, the sequence alignment tool used is Clustal Omega ([www\(.\).ebi\(.\).ac\(.\).uk](http://www.ebi.ac.uk)).

The amino acid numbers referred to herein correspond to the amino acids in SEQ ID NO: 1 and as described above, a person skilled in the art can determine equivalent amino acid positions in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 by alignment. The addition or deletion of amino acids from the variant and/or fragment of SEQ ID NO:1 could lead to a difference in the actual amino acid position of the consensus sequence in the mutated sequence, however, by lining the mutated sequence up with the reference sequence, the amino acid in an equivalent position to the corresponding amino acid in the reference sequence can be identified and hence the appropriate position for addition or substitution of the consensus sequence can be established.

In certain embodiments, a modified Sap2 protein of the invention is an isolated modified Sap2 protein. In other embodiments, a modified Sap2 protein of the invention is a recombinant modified Sap2 protein. In yet other embodiments, a modified Sap2 protein of the invention is an isolated recombinant modified Sap2 protein.

Consensus sequence

In certain embodiments, a modified Sap2 protein of the invention comprises a D/E-X-N-Z-S/T, K-D/E-X-N-Z-S/T or J-U-B-D/E-X-N-Z-S/T-J-U-B consensus sequence, wherein X and Z are independently any amino acid except proline and J, U and B independently comprise 1 to 5 naturally occurring amino acid residues. The classical 5 amino acid glycosylation consensus sequence (D/E-X-N-Z-S/T) may be extended by 1-5 other amino acid residues either side of the consensus sequence for more efficient glycosylation J-U-B-D/E-X-N-Z-S/T-J-U-B (*e.g.* G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G (SEQ ID NO: 2)). The classical 5 amino acid glycosylation consensus sequence (D/E-X-N-Z-S/T) may be extended by lysine residues for more efficient glycosylation (*e.g.* K-D/E-X-N-Z-S/T). Thus, consensus sequences in the modified Sap2 protein of the invention may comprise (or consist of) a D/E-X-N-Z-S/T consensus sequence.

In a modified Sap2 protein of the invention, the consensus sequence(s) may be selected from: D/E-X-N-Z-S/T, K-D/E-X-N-Z-S/T (SEQ ID NO: 18) or J-U-B-D/E-X-N-Z-S/T-J-U-B, wherein X is Q (glutamine) and Z is A (alanine). In the modified Sap2 protein of the invention, the consensus sequence(s) may be selected from: D/E-X-N-Z-S/T and K-D/E-X-N-Z-S/T (SEQ ID NO: 18), wherein X is Q (glutamine) and Z is A (alanine). In some embodiments, the consensus sequence is D/E-X-N-Z-S/T, wherein X is Q (glutamine) and Z is A (alanine), *e.g.* D-Q-N-A-T (SEQ ID NO: 4) also referred to as "DQNAT" (SEQ ID NO: 4). In other embodiments, the consensus sequence is K-D/E-X-N-Z-S/T (SEQ ID NO: 18), wherein X is Q (glutamine) and Z is A (alanine), *e.g.* K-D-Q-N-A-T (SEQ ID NO: 19) also referred to as "KDQNAT" (SEQ ID NO: 19). In yet other embodiments, the consensus sequence is K-D/E-X-N-Z-S/T (SEQ ID NO: 18), wherein X is Q (glutamine) and Z is A (alanine), *e.g.* K-D-Q-N-A-S (SEQ ID NO: 20) also referred to as "KDQNAS" (SEQ ID NO: 20). In a modified Sap2 protein of the invention, the consensus sequence(s) may be selected from: D/E-X-N-Z-S/T, K-D/E-X-N-Z-S/T (SEQ ID NO: 18) or J-U-B-D/E-X-N-Z-S/T-J-U-B, wherein X is Q (glutamine), Z is A (alanine), J, U and B are independently 1 to 5 amino acid residues independently selected from glycine and/or serine. In some embodiments, the consensus sequence is J-U-B-D/E-X-N-Z-S/T-J-U-B (SEQ ID NO: 6), wherein X is Q (glutamine), Z is A (alanine), each of J and B comprises 1 to 5 Glycine (G) residues and U comprises 1 to 5 serine (S) residues, *e.g.* G-S-G-G-G-D-Q-N-A-T-G-S-G-G-G (SEQ ID NO: 7) also referred to as "GSGGGDQNATGSGGG" (SEQ ID NO: 7).

In certain embodiments, the modified Sap2 protein of the invention comprises at least two D/E-X-N-Z-S/T or K-D/E-X-N-Z-S/T consensus sequences. In other embodiment, the modified Sap2 protein of the invention comprises at least three D/E-X-N-Z-S/T or K-D/E-X-N-Z-S/T consensus sequences. In yet other embodiments, the modified Sap2 protein of the invention comprises at least four D/E-X-N-Z-S/T or K-D/E-X-N-Z-S/T consensus sequences. In still other embodiments, the modified Sap2 protein of the invention comprises at least five D/E-X-N-Z-S/T or K-D/E-X-N-Z-S/T consensus sequences. In additional embodiments, the modified Sap2 protein of the invention comprises at least six D/E-X-N-Z-S/T or K-D/E-X-N-Z-S/T consensus sequences. In further embodiments, the modified Sap2 protein of the invention comprises at least seven D/E-X-N-Z-S/T or K-D/E-X-N-Z-S/T consensus sequences. In other embodiments, the modified Sap2 protein of the invention contains three to seven D/E-X-N-Z-S/T or K-D/E-X-N-Z-S/T consensus sequences. In yet other embodiments, the modified Sap2 protein of the invention contains four to seven D/E-X-N-Z-S/T or K-D/E-X-N-Z-S/T consensus sequences. In additional embodiments, the modified Sap2 protein of the invention contains five to seven D/E-X-N-Z-S/T or K-D/E-X-N-Z-S/T consensus sequences.

Introduction of such glycosylation sites can be accomplished by, *e.g.* adding new amino acids to the primary structure of the protein (*i.e.* the glycosylation sites are added, in full or in part), or by mutating existing amino acids in the protein in order to generate the glycosylation sites (*i.e.* amino acids are not added to the protein, but selected amino acids of the protein are mutated so as to form

glycosylation sites). In certain embodiments, the consensus sequence(s) are recombinantly introduced into the Sap2 amino acid sequence of SEQ ID NO: 1 or a Sap2 amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

In some embodiments, a modified Sap2 protein of the invention may further comprise a
5 "peptide tag" or "tag", *i.e.* a sequence of amino acids that allows for the isolation and/or identification of the modified Sap2 protein. For example, adding a tag to a modified Sap2 protein of the invention can be useful in the purification of that protein and, hence, the purification of conjugate (*e.g.* bioconjugate) vaccines comprising the tagged modified Sap2 protein. Exemplary tags that can be used herein include, without limitation, histidine (HIS) tags (*e.g.* hexa histidine-tag, or 6XHis-Tag),
10 FLAG-TAG, and HA tags. In certain embodiments, the tag is a hexa-histidine tag. In certain aspects, the tags used herein are removable, *e.g.* removal by chemical agents or by enzymatic means, once they are no longer needed, *e.g.* after a modified Sap2 protein of the invention has been purified. Thus, a modified Sap2 protein of the invention may further comprise a peptide tag. In certain
15 embodiments, the peptide tag is located at the C-terminus of the amino acid sequence of a modified Sap2 protein of the invention. In some aspects, the peptide tag comprises six histidine residues at the C-terminus of the amino acid sequence of a modified Sap2 protein of the invention. In certain embodiments, the present invention provides a modified Sap2 protein comprising a tag (*e.g.*, a histidine tag). In other embodiments, the present invention provides a modified Sap2 protein not comprising a tag (*e.g.*, a histidine tag), *e.g.*, a modified Sap2 protein with the histidine tag removed.
20 Thus, in particular aspects, a modified Sap2 protein of the invention comprises (or consists of): (i) amino acid residues 19-398 of SEQ ID NO: 1 or an amino acid sequence which is at least 80%, 85%, 90%, 92%, 95%, 97%, 98%, 99% or 100% identical to amino acid residues 19-398 of SEQ ID NO: 1 and (ii) a peptide tag (*e.g.* six histidine residues at the C-terminus of the amino acid sequence). In other aspects, a modified Sap2 protein of the invention comprises (or consists of) amino acid residues
25 19-398 of SEQ ID NO: 1 or an amino acid sequence which is at least 80%, 85%, 90%, 92%, 95%, 97%, 98%, 99% or 100% identical to amino acid residues 19-398 of SEQ ID NO: 1, with the peptide tag (*e.g.* histidine tag) removed.

In other embodiments, a modified Sap2 protein of the invention comprises a signal sequence which is capable of directing the modified Sap2 protein to the periplasm of a host cell (*e.g.* bacterium).
30 Signal sequences, including periplasmic signal sequences, are usually removed during translocation of the protein into, for example, the periplasm by signal peptidases (*i.e.* a mature protein is a protein from which at least the signal sequence has been removed). The signal sequence is selected from, but is not limited to, *E. coli* flagellin (FlgI) [MIKFLSALILLVTAAQA (SEQ ID NO: 21)], *E. coli* outer membrane porin A (OmpA) [MKKTAIAIAVALAGFATVAQA (SEQ ID NO: 22)], *E. coli* maltose binding protein (MalE) [MKIKTGARILALSALTMMFSASALA (SEQ ID NO: 23)], *E. coli* outer membrane porin C
35 (OmpC) [MKVKVLSLLVPALLVAGAANA (SEQ ID NO: 24)], *Erwinia carotovora* pectate lyase (PelB)

[MKYLLPTAAAGLLLLAAQPAMA (SEQ ID NO: 38)], heat labile *E. coli* enterotoxin LTIIb [MSFKKIIKAFVIMAALVSVQAHA (SEQ ID NO: 39)], *Bacillus subtilis* endoxylanase XynA [MFKFKKKFLVGLTAAFMISMFSATASA (SEQ ID NO: 40)], *E. coli* DsbA [MKKIWLALAGLVLAFSASA (SEQ ID NO: 41)], TolB [MKQALRVAFGFLILWASVLHA (SEQ ID NO: 42)] or SipA [MKMNKKVLLTSTMAASLLSVASVQAS (SEQ ID NO: 43)]. In specific embodiments, the signal sequence is from *E. coli* flagellin (FlgI) [MIKFLSALILLVTTAAQA (SEQ ID NO: 21)]. Thus, specific embodiments, the present invention provides a modified Sap2 protein, wherein the amino acid sequence further comprises a signal sequence which is capable of directing the modified Sap2 protein to the periplasm of a host cell (*e.g.* bacterium). In certain aspects, the signal sequence is FlgI. In some aspects, the FlgI comprises an amino acid sequence of SEQ ID NO: 21. In certain embodiments, the bacterial signal sequence is removed from the modified Sap2 protein of the invention after the modified Sap2 protein is transported to the periplasmic side of the inner membrane of a host cell of the invention.

In further embodiments, the present invention provides a polynucleotide encoding a modified Sap2 protein of the invention. In certain embodiments, the present invention provides a polynucleotide encoding a modified Sap2 protein of the invention, having a nucleotide sequence that encodes a polypeptide with an amino acid sequence that is at least 97%, 98%, 99% or 100% identical to SEQ ID NO: 9 or 10. In other embodiments, the present invention provides a nucleotide sequence according to SEQ ID NO: 35, or a nucleotide sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 35. In additional embodiments, the present invention provides a nucleotide sequence according to SEQ ID NO: 36, or a nucleotide sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 36. In yet other embodiments, the present invention provides a nucleotide sequence according to SEQ ID NO: 37, or a nucleotide sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 37. In certain aspects, a nucleotide sequence of the invention comprises nucleotides encoding for amino acids corresponding to one (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T, K-D/E-X-N-Z-S/T-K, and J-U-B-D/E-X-N-Z-S/T-J-U-B. In other aspects, a nucleotide sequence of the invention comprises nucleotides encoding for amino acids corresponding to one (or more) consensus sequence(s) selected from: KDQNAT (SEQ ID NO: 19), KDQNAS (SEQ ID NO: 20), DQNAT (SEQ ID NO: 4), and GSGGGDQNATGSGGG (SEQ ID NO: 7). In particular aspects, a nucleotide sequence of the invention comprises nucleotides encoding for a modified Sap2 protein of the invention comprising one or more consensus sequences, wherein the one or more consensus sequences have been added next to or substituted for one or more amino acids selected from the group consisting of: amino acid residue 19, amino acid residue 57, one or more amino acids between amino acid residues 98-102, amino acid residue 87, one or more amino acids between amino acid residues 104-108, one or more amino acids between amino acid residues 109-113, amino acid residue 220, one or more amino acids between amino acid residues 142-144, amino acid residue 198, amino acid residue 248,

amino acid residue 261, amino acid residue 268, amino acid residue 325, amino acid residue 344, one or more amino acids between amino acid residues 358-359, and amino acid residue 398 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of
5 SEQ ID NO: 1.

In additional embodiments, the present invention provides a vector comprising a polynucleotide encoding a modified Sap2 protein of the invention.

Conjugates

10 In certain embodiments, the present invention provides a conjugate comprising a modified Sap2 protein of the invention. The conjugate of the invention may be a conjugate of a modified Sap2 protein (*e.g.* chemical conjugate or bioconjugate). The conjugate of the invention may be a conjugate of a modified Sap2 protein and an antigen, *e.g.* a saccharide antigen (*i.e.* bioconjugate). In specific
15 embodiments, the present invention provides a conjugate comprising (or consisting of) a modified Sap2 protein of the invention and at least one saccharide antigen. In certain embodiments, the conjugate of the invention is a bioconjugate.

In some embodiments, the present invention provides a conjugate comprising a saccharide of the invention linked to a modified carrier protein. In certain aspects, the saccharide of the invention is linked to an asparagine residue of the modified carrier protein. Thus, in specific embodiments, the
20 present invention provides a conjugate comprising a saccharide of the invention linked to an asparagine residue of a modified carrier protein. In some embodiments, the conjugate is a bioconjugate. In certain aspects, the modified carrier protein of the invention includes, but is not limited to, Sap2, Als3, detoxified Exotoxin A of *P. aeruginosa* (EPA), CRM197, Diphtheria toxoid, tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A of *S. aureus*, clumping factor B
25 of *S. aureus*, *E. coli* FimH, *E. coli* FimHC, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* sat protein, the passenger domain of *E. coli* sat protein, *C. jejuni* AcrA, and a *C. jejuni* natural glycoprotein. In specific aspects, the modified carrier protein is a modified Sap2 protein of the invention.

30 In some embodiments, the modified Sap2 protein of the invention is linked to the at least one saccharide antigen. In certain aspects, the modified Sap2 protein of the invention is directly linked to the at least one saccharide antigen. In other aspects, the modified Sap2 protein of the invention is linked to the at least one saccharide antigen through a linker. In certain embodiments, the modified Sap2 protein of the invention is non-covalently linked to the at least one saccharide antigen. In some
35 aspects, the modified Sap2 protein of the invention is non-covalently linked to the at least one saccharide antigen via an avidin-streptavidin interaction. In other embodiments, the modified Sap2

protein of the invention is covalently linked to the at least one saccharide antigen through a chemical linkage obtainable using a chemical conjugation method (*i.e.* the conjugate is produced by chemical conjugation). The chemical conjugation method may be selected from the group consisting of carbodiimide chemistry, reductive amination, cyanylation chemistry (for example CDAP chemistry), maleimide chemistry, hydrazide chemistry, ester chemistry, and N-hydroxysuccinimide chemistry. Conjugates can be prepared by direct reductive amination methods as described in, US200710184072 (Hausdorff) US 4365170 (Jennings) and US 4673574 (Anderson). Other methods are described in EP-0-161-188, EP-208375 and EP-0-477508. The conjugation method may alternatively rely on activation of the saccharide with 1-cyano-4-dimethylamino pyridinium tetrafluoroborate (CDAP) to form a cyanate ester. Such conjugates are described in PCT published application WO 93/15760 Uniformed Services University and WO 95/08348 and WO 96/29094. See also Chu C. *et al.* Infect. Immunity, 1983 245 256.

In general the following types of chemical groups on a modified Sap2 protein can be used for coupling / conjugation:

A) Carboxyl (for instance via aspartic acid or glutamic acid). In one embodiment this group is linked to amino groups on saccharides directly or to an amino group on a linker with carbodiimide chemistry *e.g.* with EDAC.

B) Amino group (for instance via lysine). In one embodiment this group is linked to carboxyl groups on saccharides directly or to a carboxyl group on a linker with carbodiimide chemistry *e.g.* with EDAC. In another embodiment this group is linked to hydroxyl groups activated with CDAP or CNBr on saccharides directly or to such groups on a linker; to saccharides or linkers having an aldehyde group; to saccharides or linkers having a succinimide ester group.

C) Sulphydryl (for instance via cysteine). In one embodiment this group is linked to a bromo or chloro acetylated saccharide or linker with maleimide chemistry. In one embodiment this group is activated/modified with bis diazobenzidine.

D) Hydroxyl group (for instance via tyrosine). In one embodiment this group is activated/modified with bis diazobenzidine.

E) Imidazolyl group (for instance via histidine). In one embodiment this group is activated/modified with bis diazobenzidine.

F) Guanidyl group (for instance via arginine).

G) Indolyl group (for instance via tryptophan).

On a saccharide, in general the following groups can be used for a coupling: OH, COOH or NH₂. Aldehyde groups can be generated after different treatments such as: periodate, acid hydrolysis, hydrogen peroxide, etc.

Conjugates can be purified by any method known in the art for purification of a protein, for example, by chromatography (*e.g.* ion exchange, anionic exchange, affinity, and sizing column

chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. See, *e.g.*, Saraswat *et al.*, 2013, Biomed. Res. Int. ID0312709 (p. 1-18); see also the methods described in WO 2009/104074. The actual conditions used to purify a particular conjugate will depend, in past, on the synthesis strategy (*e.g.*, synthetic production vs. recombinant production) and on factors such as net charge, hydrophobicity, and/or hydrophilicity of the bioconjugate.

In some embodiments, the amino acid residue on the modified Sap2 protein of the invention to which the at least one saccharide antigen is linked includes, without limitation, Ala, Arg, Asp, Cys, Gly, Glu, Gln, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, and Val. In additional embodiments, the amino acid is: an amino acid comprising a terminal amine group, a lysine, an arginine, a glutamic acid, an aspartic acid, a cysteine, a tyrosine, a histidine or a tryptophan. In certain embodiments, the amino acid residue on the modified Sap2 protein of the invention to which the at least one saccharide antigen is linked is not an asparagine residue and in this case, the conjugate is typically produced by chemical conjugation. In other embodiments, the at least one saccharide antigen is linked to an amino acid on a modified Sap2 protein of the invention selected from asparagine, aspartic acid, glutamic acid, lysine, cysteine, tyrosine, histidine, arginine or tryptophan (*e.g.* asparagine) and in the case of asparagine, the conjugate may be a bioconjugate (for example an enzymatic conjugation using a oligosaccharyl transferase such as PglB). In specific embodiments, the amino acid residue on a modified Sap2 protein of the invention to which the at least one saccharide antigen is linked is an asparagine residue. In specific embodiments, the amino acid residue on the modified Sap2 protein to which the at least one saccharide antigen is linked is part of the consensus sequence, *e.g.* the asparagine in D/E-X-N-Z-S/T, K-D/E-X-N-Z-S/T-K or J-U-B-D/E-X-N-Z-S/T-J-U-B consensus sequence.

In certain embodiments, the conjugate of the invention is a conjugate of a recombinant modified Sap2 protein (*e.g.* chemical conjugate or bioconjugate). In other embodiments, the conjugate of the invention is a conjugate of an isolated recombinant modified Sap2 protein and a recombinant antigen, *e.g.* recombinant saccharide antigen (*i.e.* bioconjugate).

In certain embodiments, the modified Sap2 protein of the invention is linked to the at least one saccharide antigen at one or more amino acid residues on the modified Sap2 protein. In certain aspects, the one or more residues include, without limitations, one or more asparagine residues, one or more aspartic acid residues, one or more glutamic acid residues, one or more lysine residues, one or more cysteine residues, one or more tyrosine residues, one or more histidine residues, one or more arginine residues, one or more tryptophan residues, one or more serine residues, and one or more threonine residues. In specific embodiments, the modified Sap2 protein of the invention is linked to the at least one saccharide antigen at one or more asparagine residues on the modified Sap2 protein of the invention. In certain aspects, the at least one saccharide antigen is linked to at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least

nine, or at least ten asparagine residues of a modified Sap2 protein of the invention. In some aspects, the at least one saccharide antigen is linked to one, two, three, four, five, six, seven, eight, nine, or ten asparagine residues of a modified Sap2 protein of the invention. In certain embodiments, the at least one saccharide antigen is linked to at least three asparagine residues of a modified Sap2 protein of the invention. In specific embodiments, the at least one saccharide antigen is linked to four asparagine residues of a modified Sap2 protein of the invention. In certain aspects, the modified Sap2 protein of the invention comprises (or consists of) an amino acid sequence of SEQ ID NO: 9, and the four asparagine residues include, but are not limited to, positions 45, 94, 215, and 415 of SEQ ID NO: 9. In certain embodiments, the present invention provides a modified Sap2 protein of *Candida albicans* comprising (or consisting of): (1) an amino acid sequence of SEQ ID NO: 9 or SEQ ID NO: 10; and (2) at least one saccharide antigen of *Candida*, wherein the at least one saccharide antigen is a β -1,2 mannan polymer consisting of at least five consecutive β -1,2 linked mannose molecules, and wherein the at least one saccharide chain is linked to at least one of four asparagine residues at positions 45, 94, 215, and 415 of SEQ ID NO: 9 or at least one of four asparagine residues at positions 6, 55, 176, and 376 of SEQ ID NO: 10.

Antigens

The present invention provides conjugates (e.g., bioconjugates) wherein a modified Sap2 protein of the invention may be linked to a number of different antigens. In certain embodiments, the modified Sap2 protein is linked to at least one antigen which is a saccharide antigen. In certain aspects, the antigen comprises at least one saccharide antigen. In some embodiments, the at least one saccharide antigen is a fungal polysaccharide, a yeast polysaccharide or a mammalian polysaccharide. Polysaccharides comprise 2 or more monosaccharides, typically greater than 6, 8 or 10 monosaccharides.

In certain aspects, the at least one saccharide antigen in a conjugate (e.g. bioconjugate) of the invention includes, but is not limited to, *Citrobacter freundii* O antigen, *Agrobacterium* sp. exopolysaccharide, O antigens of *E. coli*, *Salmonella* sp. O antigens, *Pseudomonas* sp., *Klebsiella* sp. O antigens, *Acinetobacter* O antigens, *Chlamydia trachomatis* O antigens, *Vibrio cholera* O antigens, *Listeria* sp. O antigens, *Legionella pneumophila* serotypes 1 to 15 O antigens, *Bordetella parapertussis* O antigens, *Burkholderia mallei* and *pseudomallei* O antigens, *Francisella tularensis* O antigens, *Campylobacter* sp. O antigens, capsular polysaccharides of *Clostridium difficile*, *Staphylococcus aureus* type 5 and 8, *Streptococcus pyogenes*, *E. coli*, *Streptococcus agalacticae*, *Neisseria meningitidis*, *Candida* sp., *Candida albicans*, *Haemophilus influenza*, *Enterococcus faecalis* capsular polysaccharides type I-V, and other surface polysaccharide structures, e.g. the *Borrelia burgdorferi* glycolipids, *Neisseria meningitidis* pilin O glycan and lipooligosaccharide (LOS), *Haemophilus influenza* LOS,

Leishmania major lipophosphoglycan, tumor associated carbohydrate antigens, malaria glycosyl phosphatidylinositol, and *mycobacterium tuberculosis* arabinomannan.

In certain embodiments, the at least one saccharide antigen is an O-antigen *e.g.* from a Gram-negative bacterium. In certain embodiments, the at least one saccharide antigen is an O-antigen from *Salmonella* species, *Shigella* species, *Pseudomonas* species or *Klebsiella* species. In other embodiments, the at least one saccharide antigen is an O-antigen from *Shigella* species, *Pseudomonas* species or *Klebsiella* species (*e.g.* *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, or *Klebsiella pneumoniae*). In other embodiments, the at least one saccharide antigen is an O-antigen from *Shigella dysenteriae*, *Shigella flexneri* or *Shigella sonnei*. For example, the antigen may be an O-antigen from *S. dysenteriae* type 1, *S. sonnei*, and *S. flexneri* type 6, and *S. flexneri* 2a and 3a 0 (Dmitriev, B.A., *et al.* Somatic Antigens of *Shigella* Eur J. Biochem, 1979. 98: p. 8; Liu et al Structure and genetics of *Shigella* O antigens FEMS Microbiology Review, 2008. 32: p. 27). In other embodiments, the at least one saccharide antigen is an O-antigen from *Pseudomonas aeruginosa*. For example, the antigen may be an O-antigen from *Pseudomonas aeruginosa* serotypes 1-20 (Raymond *et al.*, J Bacteriol. 2002 184(13):3614-22). In yet other embodiments, the at least one saccharide antigen is an O-antigen from *Klebsiella pneumoniae*.

In certain embodiments, the at least one saccharide antigen is a capsular polysaccharide from *Neisseria meningitidis* serogroup A (MenA), *N. meningitidis* serogroup C (MenC), *N. meningitidis* serogroup Y (MenY), *N. meningitidis* serogroup W (MenW), *H. influenzae* type b (Hib), Group B *Streptococcus* (GBS), *Streptococcus pneumoniae*, or *Staphylococcus aureus*. In other embodiments, the at least one saccharide antigen is a capsular polysaccharide from *Streptococcus* species or *Staphylococcus* species. (*e.g.* *Streptococcus pneumoniae* or *Staphylococcus aureus*). In additional embodiments, the at least one saccharide antigen is a capsular polysaccharide from *Staphylococcus aureus*. For example, the at least one saccharide antigen may be a capsular polysaccharide from *Staphylococcus aureus* type 5 and 8. In other embodiments, the at least one saccharide antigen is a capsular polysaccharide from *Streptococcus pneumoniae*.

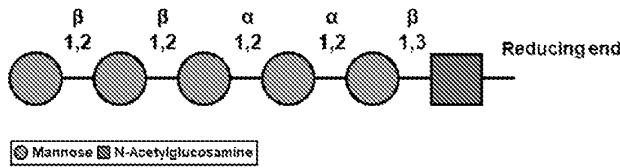
In additional embodiments, the at least one saccharide antigen is a fungal saccharide antigen. In certain aspects, the fungi is *Candida* species. Thus, in certain aspects, the at least one saccharide antigen is a saccharide antigen of *Candida* species. In some aspects, the *Candida* species includes, but is not limited to, *Candida albicans*, *Candida auris*, *Candida guilliermondii*, *Candida lusitanae* and *Candida tropicalis*, *Candida glabrata*, *Candida krusei*, and *Candida parapsilosis*. In specific embodiments, the *Candida* species is *Candida albicans*. Thus, in specific embodiments, the at least one saccharide antigen is a saccharide antigen of *Candida albicans*.

In certain embodiments, the at least one saccharide antigen is a β -1,2 mannan polymer. In certain aspects, the at least one saccharide antigen is a β -1,2 mannan polymer of *Candida albicans*. Mannans form the outermost layers of the cell wall and they participate in *Candida* cell adherence and

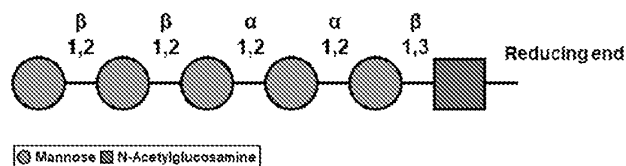
immune system evasion. Mannans are highly complex and branched structures which are linked to secreted proteins. β -1,2 mannans are a part of *Candida*'s mannans present at the non-reducing end. Structures comprising at least two consecutive β -1,2 mannoses are exclusive to the *Candida* genus and are conserved in several *Candida* species (e.g., including *C. albicans*, *C. auris*, *C. guilliermondii*,
5 *C. lusitanae*, *C. tropicalis*, and excluding *C. glabrata*) (Gow NA & Hube B; 2012; Curr Opin Microbiol; Miyakawa, Y et al, Infect. Immun 1992; Shibata N et al, Proc Jpn Acad, Ser B, 2012; Rudkin FM et al, Nat Commun 2018; Morad HO et al, *Front Microbiol*, 2018; Han et al, J. Infect. Dis, 1999; Xin et al, PNAS 2008). In certain embodiments, the β -1,2 mannan polymer of the invention is a naturally-occurring mannan. In other embodiments, the β -1,2 mannan polymer of the invention is a
10 recombinantly produced mannan.

In certain embodiments, the at least one saccharide antigen comprises a β -1,2 mannan polymer. In some aspects, the β -1,2 mannan polymer comprises at least two β -1,2 linked mannose molecules. In additional aspects, the β -1,2 mannan polymer comprises two to fifty, two to forty, two to thirty, two to twenty, two to ten, two to nine, two to eight, two to seven, two to six, two to five,
15 two to four, two to three, two, three to fifty, three to forty, three to thirty, three to twenty, three to ten, three to nine, three to eight, three to seven, three to six, three to five, three to four, three, four to fifty, four to forty, four to thirty, four to twenty, four to ten, four to nine, four to eight, four to seven, four to six, four to five, four, five to fifty, five to forty, five to thirty, five to twenty, five to ten, five to nine, five to eight, five to seven, five to six, or five, six to forty, six to thirty, six to twenty, six
20 to ten, six to nine, six to eight, six to seven, six, seven to fifty, seven to forty, seven to thirty, seven to twenty, seven to ten, seven to nine, seven to eight, or seven β -1,2 linked mannose molecules. In some aspects, the β -1,2 mannan polymer comprises at least 2, at least 3, at least 4, or at least 5 β -1,2 linked mannose molecules. In additional aspects, the β -1,2 mannan polymer comprises at least 2, at least 3, at least 4, or at least 5 consecutive β -1,2 linked mannose molecules.

25 In certain embodiments, the present invention provides a saccharide that is a β -1,2 mannan polymer comprising the structure:



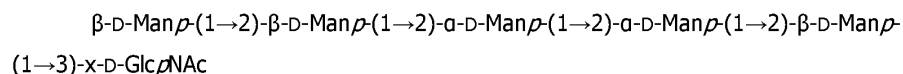
In some embodiments, the present invention provides a saccharide having the structure:



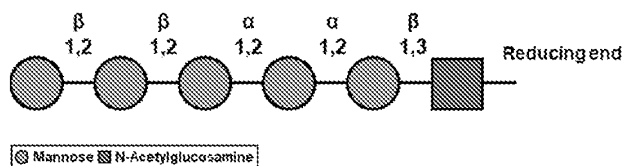
In certain embodiments, the present invention provides a saccharide comprising the structure:



- 5 In additional embodiments, the present invention provides a saccharide comprising the structure:



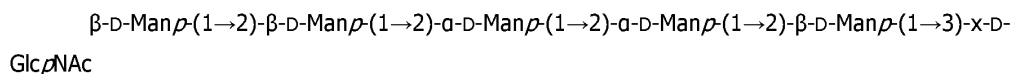
- 10 In certain embodiments, the at least one saccharide antigen comprises a glucan having the structure:



In other embodiments, the at least one saccharide antigen comprises a glucan having the structure:



- 15 In yet other embodiments, the at least one saccharide antigen comprises a glucan having the structure:



- 20 In other embodiments, the at least one saccharide antigen is a β -1,3 glucan polymer. Thus, in certain aspects, the at least one saccharide antigen is a β -1,3 glucan polymer of *Candida albicans*. β -1,3 glucans are widespread in nature. They are present in most yeasts and widespread in fungi and plants. *Candida* β -1,3 glucans consist of long linear polymers with occasional β -1,6 branchings. They confer strength and shape to the cell wall. In certain embodiments, the β -1,3 glucan polymer

of the invention is a naturally-occurring glucan. In other embodiments, the β -1,3 glucan polymer of the invention is a recombinantly produced glucan.

In certain embodiments, the at least one saccharide antigen comprises a β -1,3 glucan polymer. In some aspects, the β -1,3 glucan polymer of the invention comprises at least four β -1,3 linked glucose molecules. In additional aspects, the β -1,3 glucan polymer comprises four to hundred, 5 four to fifty, four to forty, four to thirty five, four to thirty, four to twenty five, four to twenty, four to ten, four to nine, four to eight, four to seven, four to six, four to five, four, five to hundred, five to fifty, five to forty, five to thirty five, five to thirty, five to twenty five, five to twenty, five to ten, five to nine, five to eight, five to seven, five to six, five, six to hundred, six to fifty, six to forty, six to thirty 10 five, six to thirty, six to twenty five, six to twenty, six to ten, six to nine, six to eight, six to seven, six, seven to hundred, seven to fifty, seven to forty, seven to thirty five, seven to thirty, seven to twenty five, seven to twenty, seven to ten, seven to nine, seven to eight, or seven β -1,3 linked glucose molecules. In some aspects, the β -1,3 glucan polymer comprises at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, or at least 11 consecutive β -1,3 linked glucose molecules. In specific embodiments, the β -1,3 glucan polymer comprises at least 11 β -1,3 linked glucose 15 molecules. In particular embodiments, the β -1,3 glucan polymer comprises at least 11 consecutive β -1,3 linked glucose molecules.

Thus, in certain embodiments, the present invention provides a modified Sap2 protein of *Candida albicans* comprising (or consisting of): (1) an amino acid sequence of SEQ ID NO: 9 or SEQ 20 ID NO: 10; and (2) at least one saccharide antigen of *Candida*, wherein the at least one saccharide antigen is a β -1,2 mannan polymer comprising (or consisting of) at least five consecutive β -1,2 linked mannose molecules, and wherein the at least one saccharide antigen is linked to at least one of four asparagine residues at positions 45, 94, 215, and 415 of SEQ ID NO: 9 or at least one of four asparagine residues at positions 6, 55, 176, and 376 of SEQ ID NO: 10.

25

Host cell

In certain embodiments, the present invention provides host cells comprising a polynucleotide sequence that encodes a modified Sap2 protein of the invention. In certain embodiments, the present invention provides host cells that can be used to produce the bioconjugates of the invention. Host 30 cells of the invention include, without limitation, archaea, prokaryotic host cells, and eukaryotic host cells. In certain embodiments, the host cell is a non-human host cell. In certain embodiments, the host cell is a eukaryotic host cell. In some aspects, the eukaryotic host cell includes, but is not limited to, a yeast cell, an insect cell and a mammalian cell. In additional embodiments, the host cell is a prokaryotic host cell. In other aspects, the prokaryotic host cell is a bacterial cell. In some aspects, 35 the bacteria is a gram positive bacteria. In other aspects, the bacteria is a gram negative bacteria. In certain embodiments, the bacteria includes, but is not limited to, an *Escherichia* species, a *Shigella*

species, *Klebsiella* species, a *Xhantomonas* species, a *Salmonella* species, a *Yersinia* species, a *Lactococcus* species, a *Lactobacillus* species, a *Pseudomonas* species, a *Corynebacterium* species, a *Streptomyces* species, a *Streptococcus* species, a *Staphylococcus* species, a *Bacillus* species, and a *Clostridium* species. In some aspects, the bacteria is an *E. coli* species. In specific aspects, the
5 bacteria is *E. coli*. Thus, exemplary host cells include, but are not limited to, an *Escherichia* species, a *Shigella* species, *Klebsiella* species, a *Xhantomonas* species, a *Salmonella* species, a *Yersinia* species, a *Lactococcus* species, a *Lactobacillus* species, a *Pseudomonas* species, a *Corynebacterium* species, a *Streptomyces* species, a *Streptococcus* species, a *Staphylococcus* species, a *Bacillus* species, and a *Clostridium* species. In some aspects, the host cell is an *E. coli* species. In specific embodiments, the
10 host cell is *E. coli*.

Host cells of the invention may be modified to delete or modify genes in the host cell genetic background (genome) that compete or interfere with the synthesis of the polysaccharide antigen of interest (*e.g.* compete or interfere with one or more heterologous polysaccharide synthesis genes that are recombinantly introduced into the host cell). These genes can be deleted or modified in the host
15 cell background (genome) in a manner that makes them inactive/dysfunctional (*i.e.* the host cell nucleotide sequences that are deleted/modified do not encode a functional protein or do not encode a protein whatsoever). In certain embodiments, when nucleotide sequences are deleted from the genome of the host cells of the invention, they are replaced by a desirable sequence, *e.g.* a sequence that is useful for glycoprotein production. Exemplary genes that can be deleted in host cells (and, in
20 some cases, replaced with other desired nucleotide sequences) include genes of host cells involved in glycolipid biosynthesis, such as *waaL* (see, *e.g.* Feldman et al. 2005, PNAS USA 102:3016-3021), the O antigen cluster (*rfb* or *wb*), enterobacterial common antigen cluster (*wec*), the lipid A core biosynthesis cluster (*waa*), galactose cluster (*gal*), arabinose cluster (*ara*), colonic acid cluster (*wc*), capsular polysaccharide cluster, undecaprenol-pyrophosphate biosynthesis genes (*e.g.* *uppS* (Undecaprenyl pyrophosphate synthase), *uppP* (Undecaprenyl diphosphatase)), Und-P recycling
25 genes, metabolic enzymes involved in nucleotide activated sugar biosynthesis, enterobacterial common antigen cluster, and prophage O antigen modification clusters like the *gtrABS* cluster. In some embodiments, one or more of the *waaL* gene, *gtrA* gene, *gtrB* gene, *gtrS* gene, or a gene or genes from the *wec* cluster or a gene, or a gene or genes from the colonic acid cluster (*wc*), or a gene
30 or genes from the *rfb* gene cluster are deleted or functionally inactivated from the genome of a prokaryotic host cell of the invention. In other embodiments, one or more of the *waaL* gene, *gtrA* gene, *gtrB* gene, *gtrS* gene, or a gene or genes from the *wec* cluster or a gene or genes from the *rfb* gene cluster are deleted or functionally inactivated from the genome of a prokaryotic host cell of the invention. In additional embodiments, the host cell of the invention is *E. coli*, wherein the native enterobacterial common antigen cluster (ECA, *wec*) with the exception of *wecA*, the colonic acid cluster (*wca*), and the O16-antigen cluster have been deleted. In further embodiments, the native
35

lipopolysaccharide O-antigen ligase *waaL* may be deleted from the host cell of the invention. In other embodiments, the native *gtrA* gene, *gtrB* gene and *gtrS* gene, may be deleted from the host cell of the invention.

The host cells of the present invention are engineered to comprise heterologous nucleotide sequences. In certain aspects, the host cells of the present invention are engineered to comprise a nucleotide sequence that encodes a modified Sap2 protein of the invention, optionally within a plasmid. In other aspects, the host cells of the invention also comprise one or more nucleotide sequences comprising polysaccharide synthesis genes. Thus, host cells of the invention can produce a bioconjugate comprising an antigen, for example a saccharide antigen (*e.g.* a fungal, bacterial, yeast or mammalian polysaccharide antigen) which is linked to a modified Sap2 protein of the invention. In certain embodiments, one or more heterologous nucleotide sequences encode for the polysaccharide synthesis proteins to produce the fungal polysaccharide antigen, bacterial polysaccharide antigen, yeast polysaccharide antigen or mammalian polysaccharide antigen.

Thus, in certain embodiments, the present invention provides a host cell comprising: (1) one or more polynucleotide sequences that encode one or more heterologous glycosyltransferases; (2) a polynucleotide sequence that encodes a heterologous oligosaccharyl transferase; (3) a polynucleotide sequence that encodes a modified Sap2 protein of the invention; and, optionally, (4) a polynucleotide sequence that encodes a polymerase.

In other embodiments, the present invention provides a host cell comprising: (1) one or more polynucleotide sequences that encode one or more heterologous glycosyltransferases; (2) a polynucleotide sequence that encodes a heterologous oligosaccharyl transferase; (3) a polynucleotide sequence that encodes a modified Sap2 protein of the invention (optionally a polynucleotide sequence of the invention); and, optionally, (4) a polynucleotide sequence that encodes a polymerase.

In some aspects, the one or more heterologous glycosyltransferases includes, without limitation, WbaD, WbaC, and WbaB. In certain embodiments, the WbaD, WbaC, and WbaB are from *Citrobacter*. In specific aspects, the *Citrobacter* is *Citrobacter freundii*. In certain aspects, the *Citrobacter freundii* is *Citrobacter freundii* P079F I. In certain embodiments, the one or more heterologous glycosyltransferases comprises WbaD, WbaC, and WbaB from *Citrobacter*, optionally from *Citrobacter freundii*, optionally *Citrobacter freundii* P079F I. In certain embodiments, the one or more heterologous glycosyltransferase(s) has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of WbaD of *Citrobacter freundii* P079F I comprising SEQ ID NO: 11. In other embodiments, the one of more heterologous glycosyltransferase(s) has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of WbaC of *Citrobacter freundii* P079F I comprising SEQ ID NO: 12. In still other embodiments, the one of more heterologous glycosyltransferase(s) has an amino acid sequence at least 80%, 90%, 95%, 98%,

or 99% identical to the amino acid sequence of WbaB of *Citrobacter freundii* P079F I comprising SEQ ID NO: 13.

In additional embodiments, the present invention provides a host cell comprising:

- i. a nucleotide sequence encoding one or more heterologous glycosyltransferase(s) capable of synthesizing a β -1,2 mannan polymer;
- ii. a nucleotide sequence encoding a second heterologous glycosyltransferase which is eukaryotic and is capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain;
- iii. a nucleotide sequence encoding a heterologous oligosaccharyl transferase; and
- iv. optionally, a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.

In yet other embodiments, the present invention provides a host cell comprising:

- i. a nucleotide sequence encoding one or more heterologous glycosyltransferase(s) capable of synthesizing a β -1,2 mannan polymer;
- ii. a nucleotide sequence encoding a second heterologous glycosyltransferase which is eukaryotic and is capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain;
- iii. a nucleotide sequence encoding a heterologous oligosaccharyl transferase; and
- iv. optionally, a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline, optionally a polynucleotide sequence of the invention.

In certain aspects, the modified carrier protein of the invention includes, but is not limited to, Sap2, Als3, detoxified Exotoxin A of *P. aeruginosa* (EPA), CRM197, Diphtheria toxoid, tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A of *S. aureus*, clumping factor B of *S. aureus*, *E. coli* FimH, *E. coli* FimHC, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* sat protein, the passenger domain of *E. coli* sat protein, *C. jejuni* AcrA, and a *C. jejuni* natural glycoprotein. In specific aspects, the modified carrier protein is a modified Sap2 protein of the invention.

In some embodiments, the one or more heterologous glycosyltransferases capable of synthesizing a β -1,2 mannan polymer includes, without limitation, WbaD, WbaC, and WbaB. In certain embodiments, the WbaD, WbaC, and WbaB are from *Citrobacter*. In specific aspects, the *Citrobacter* *Citrobacter freundii*. In certain aspects, the *Citrobacter freundii* is *Citrobacter freundii* P079F I. In certain embodiments, the one or more heterologous glycosyltransferases comprises WbaD, WbaC,

and WbaB from *Citrobacter*, optionally from *Citrobacter freundii*, optionally *Citrobacter freundii* P079F I.

In certain aspects, the one or more heterologous glycosyltransferases capable of synthesizing a β -1,2 mannan polymer has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of WbaD of *Citrobacter freundii* P079F I. In some aspects, the one or more heterologous glycosyltransferases has an amino acid sequence identical to the amino acid sequence of WbaD of *Citrobacter freundii* P079F I. In certain embodiments, WbaD of *Citrobacter freundii* P079F I comprises the amino acid sequence of SEQ ID NO: 11. Thus, in certain aspects, the one or more heterologous glycosyltransferases has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to SEQ ID NO: 11. In other aspects, the one or more heterologous glycosyltransferases has an amino acid sequence identical to SEQ ID NO: 11:

MSNNPKKKILVLTTRYPPFVIGGDKLRIYKICQELSKYYDLTLLTLIDDIQDLTIPHDEEVFKYVHKIYLPK
IKSFFNVLLALPTSTPLQVAYYKSDNFKQKLNELLPAYDATLSHLIRVGHYAKDVNGVNFLEMTDAISLNYKRVRE
IKTLKSFKSFIFSLEQKRLERYERSIASSFDLTTFISAVDKNFLYPEERTDVIVSGNGVDTNFLQFKNRHIKSQEPV
LIFIGNMLSLQNMDAVTFFAKKILPLLNEKGNFIFKVIKISEKSRRILSAIPDVIVTGTVDNILETASDGHIGICSM
RLGAGVQNKVLEYMALGMPCVTTTVGFEGIGARDGNDIADSPREYVTAIEKLVNDSNYFSSIAINARNFVVSQ
YSWEMQLSTFVASVNKLLK

In certain embodiments, host cells of the present invention comprise a nucleotide sequence encoding WbaD, optionally WbaD of *Citrobacter freundii* P079F I, optionally a nucleotide sequence encoding WbaD of *Citrobacter freundii* P079F I comprising an amino acid sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 11, optionally within a plasmid.

In additional embodiments, host cells of the present invention comprise a nucleotide sequence encoding WbaD, optionally WbaD of *Citrobacter freundii* P079F I, optionally a nucleotide sequence encoding WbaD of *Citrobacter freundii* P079F I comprising a nucleotide sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 25:

ATGAGCAATAACCCCAAGAAAAAATTCTTGTTCTAACGCCTCGCTATCCATTTCTGTAATCGGTG
GTGATAAACTGAGAATTTATAAAATATGTCAGGAACCTTCAAATACTATGATCTTACATTATTAACCTTAATT
GATGATATTCAGGATTTAACCATACCTCATGATGAGGAAGTTTTAAATACGTACATAAGATATATCTCCCTA
AAATAAAGTCTTTTTTAAATGTACTTTTGGCATTACCTACTTCTACGCCATTGCAGGTTGCATATTATAAATCT
GATAATTTTAAGCAAAACTGAATGAATTGTTGCCTGCCTATGATGCAACACTTTCTCATTTGATCAGAGTTG
GGCATTATGCCAAGGATGTGAATGGTGTGAATTTCTTGAAATGACAGATGCCATATCTCTTAACACAAAAG
AGTTAGGGAAATTAACCACTGAAAAGTTTTAAATCCTTCATTTTTCTTTAGAGCAAAAAGATTGGAGCGC
TATGAACGTTCAATTGCAAGTTCTTTGATTTAACGACATTTATTTCTGCGGTGGATAAGAATTTTTATATC
CAGAGGAAAGAACTGATGTAATTGTATCAGGAATGGTGTAGATACAAATTTCTTCAGTTTAAAAATCGCCA
TATTAATCACAAGAACCTGTTGTTTTGATTTTTATCGGGAATATGCTCTCATTGCAAAATATGGATGCTGTC

ACATTCTTTGCGAAAAAATACTACCACTTCTTAACGAAAAAGGTAATTCATATTTAAAGTTATCGGTAAAA
 TTTCTGAGAAGAGTAGAAGGATTCTGTCCGCAATCCCTGATGTGATAGTCACTGGTACTGTTGATAATATTTT
 AGAAACTGCATCTGATGGGCACATTGGTATTTGCTCAATGAGATTGGGTGCTGGAGTACAGAATAAAGTATT
 AGAATATATGGCCTTGGGTATGCCCTGTGTAACGACTGTGGGGTTTGAAGGCATTGGGGCTAGGGATG
 5 GTAACGACATAGTTATTGCTGACTCACCCGAGAATATGTAACGCAATTGAAAAATTAGTCAATGACAGCAA
 TTATTTTTCTTCTATAGCAATCAATGCAAGGAATTTTGTGTGTCCCAATACTCTTGGGAAATGCAGTTGTGCG
 ACATTTGTTGCTTCTGTAAATAAGCTCCTAAAATAA

In some aspects, the one or more heterologous glycosyltransferases capable of synthesizing
 a β -1,2 mannan polymer has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical
 10 to the amino acid sequence of WbaC of *Citrobacter freundii* P079F I. In some aspects, the one or
 more heterologous glycosyltransferases has an amino acid sequence identical to the amino acid
 sequence of WbaC of *Citrobacter freundii* P079F I. In certain embodiments, WbaC of *Citrobacter*
freundii P079F I comprises the amino acid sequence of SEQ ID NO: 12. Thus, in certain aspects, the
 one or more heterologous glycosyltransferases has an amino acid sequence at least 80%, 90%, 95%,
 15 98%, or 99% identical to SEQ ID NO: 12. In other aspects, the one or more heterologous
 glycosyltransferases has an amino acid sequence identical to SEQ ID NO: 12:

MRIVVNNFFYGVLRGIPYITSELVAKLREEGVEVKELRCPKFLYGLPTWIHNFLFIYDQIITPLFGLFHK
 TKYNIYPYNSLSLIDLFTNKPIVIIHDFISLNKNKNKINISACYVKFCILSSSNRIRNVILISNTTAKIANKLSLFSNARQI
 LLPNTFFSFKSLSDGVQKEDHGFLLLVSGMGKNKDIDAALYFSIPIEYRIPLKILGCGGGRDLLKAKIHGRDEFN
 20 TIEILKQIPLEDVVKLYAHCKFVWAHSLAEGYGRALAEGKISGKNILCTRIPAFIEQNCSNVFYNDTESFQEHYFN
 LIENTPVVDVCELKEHIKFSEELKKIYEQ

In certain embodiments, host cells of the present invention comprise a nucleotide sequence
 encoding WbaC, optionally WbaC of *Citrobacter freundii* P079F I, optionally a nucleotide sequence
 encoding WbaC of *Citrobacter freundii* P079F I comprising an amino acid sequence at least 80%,
 25 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 12, optionally within a
 plasmid.

In additional embodiments, host cells of the present invention comprise a nucleotide sequence
 encoding WbaC, optionally WbaC of *Citrobacter freundii* P079F I, optionally a nucleotide sequence
 encoding WbaC of *Citrobacter freundii* P079F I comprising a nucleotide sequence at least 80%, 85%,
 30 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 26:

ATGAGAATTGTAGTAAATAATTTCTTTTATGGTGTGCTAAAACGTGGAATCCCTATTTATACTTCTG
 AGTTAGTCGCTAAATTAAGAGAGGAAGGAGTTGAGGTTAAAGAACTAAGATGTCCTAAGTTTCTTTACGGCT
 TACCTACTTGGATTACATAATTTCTTTTATTATTTACGATCAAATTATTACTCCTCTATTTGGTCTTTTCAT
 AAAACAAAATATAACATTTATCCATACAACAGTTTATCCTTAATAGACTTGTTCACTAATAAGCCTATAGTTAT
 35 TATTATGATTTTATTAGTCTAAATAAGAATAAAAAAATATATCTGCATGTTATGTAAATTTTGATTCTTT
 CTAGTTCAAATAGAAATTAGAAATGTAATTTTAACTCTCAACACGACAGCGAAAATAGCGAATAAATTATCATT

ATTTTCAAATGCAAGGCAAATATTATTGCCAAATACTTTCTTTTCATTTAAATCATTGAGTGATGGTGTCCAA
 AAGGAAGATCATGGTTTTTTATTATTAGTTTCAGGAATGGGGAAAAATAAGGACATTGATGCGGCTTTAGAA
 CTTTACTTTTCAATTCCTATTGAATATCGAATCCGCTAAAAATTTTAGGATGCGGGGGGGGAGGGATTTA
 CTGAAAGCAAAAATTCATGGACGAGATGAATTCAATACTATTGAAATTTTAAACAAATCCCATTAGAGGATG
 5 TGGTTAAACTATACGCGCATTGCAAGTTTGTGGGCCCACTCTTCTGCTGAAGTTACGGAAGGGCGCTGG
 CGGAAGGTAAAGATCTCAGGGAAAAATATTTTGTGCACTAGAATTCCTGCATTTATAGAGCAAAATTGTTCTAA
 TGTGTTTTATTATAACGATACTGAAAGCTTCCAAGAGCATTATTTAATTTGATTGAGAATACACCGGTTGTC
 GACGTGTGTGAGTTAAAGAACATATAAAATTCAGCGAAGAGTTAAAGAAAATTTATGAGCAATAA

In other aspects, the one or more heterologous glycosyltransferases capable of synthesizing
 10 a β -1,2 mannan polymer has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical
 to the amino acid sequence of WbaB of *Citrobacter freundii* P079F I. In some aspects, the one or
 more heterologous glycosyltransferases has an amino acid sequence identical to the amino acid
 sequence of WbaB of *Citrobacter freundii* P079F I. In certain embodiments, WbaB of *Citrobacter*
freundii P079F I comprises the amino acid sequence of SEQ ID NO: 13. Thus, in certain aspects, the
 15 one or more heterologous glycosyltransferases has an amino acid sequence at least 80%, 90%, 95%,
 98%, or 99% identical to SEQ ID NO: 13. In other aspects, the one or more heterologous
 glycosyltransferases has an amino acid sequence identical to SEQ ID NO: 13:

MIVFFTGSYPSPKCGVGDYLYKLIANILPHHSNVKVIKNSLLEFIFYAISNRKAIRHVNIQYPTIGYASNYL
 SAFKPHFVTLIARFLGIRVSITLHEFTSLSSKARFFANLFKIANNIVATSEYEHENLVKFGFNKDRVIVIPIGSNIKES
 20 DVKDKTIDFINFGIISPGKIEDFLYVIEKIRVDYETLKVVLAGYIPDNSEYADKIIAQAKQLNIEFKPNQTEDELSIL
 VGESKRAILPYSDGISERRGTALAAMINKCVVYSYAGNSSEAFNTICMLARDRELYNKLIDALHTDSADDCFIK
 AYEYALARHWDKVSEKYLRFYENCSK

In certain embodiments, host cells of the present invention comprise a nucleotide sequence
 encoding WbaB, optionally WbaB of *Citrobacter freundii* P079F I, optionally a nucleotide sequence
 25 encoding WbaB of *Citrobacter freundii* P079F I comprising an amino acid sequence at least 80%, 85%,
 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 13, optionally within a plasmid.

In additional embodiments, host cells of the present invention comprise a nucleotide sequence
 encoding WbB, optionally WbaB of *Citrobacter freundii* P079F I, optionally a nucleotide sequence
 encoding WbaB of *Citrobacter freundii* P079F I comprising a nucleotide sequence at least 80%, 85%,
 30 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 27:

ATGATAGTTTTTTTTACAGGCTCATATCCTCCGAGTAAGTGTGGTGTGGTGATTATTTGTATAAAT
 TAATAGCTAATATTTTACCACACCATTCAAATGTTAAAGTTATAAAAACTCATTGCTTGAATTTATTTTTAT
 GCCATTTCAAATAGGAAGGCCATAAGACATGTAATATACAATATCCGACAATTGGCTATGCTAGTAATTATT
 TAAGTGCAATTTAAACCTCATTTTGTGACACTATAGCTAGATTTTTGGGGATTAGAGTTTCAATAACTCTTCA
 35 TGAGTTTACTAGCCTTTCTAGTAAGGCGCGGTTTTTGTCTAATCTTTTAAAGATTGCTAATAATATCGTTGCT
 ACATCTGAGTATGAACATGAGAATCTAGTTAAATTTGGTTTTAATAAAGATAGGGTAATAGTTATTCCAATTG

GATCGAATATAAAAGAGTCCGATGTGAAAGACAAAACCTATAGACTTTATTAACCTTTGGAATTATATCACCAGG
 AAAAGGTATTGAGGATTTCTGTATGTTATTGAAAAATCAGAGTAGATTATGAACTTTGAAAGTTGTTTTA
 GCTGGCTATATACCTGACAATAGTGAGTACGCTGATAAAATCATTGCGCAAGCCAAACAGCTTAATATTGAAT
 TCAAACCTAACCAAACTGAAGATGAACTGTCTATCTTAGTTGGGGAATCCAAAAGAGCAATATTGCCTTATAG
 5 TGATGGTATTTTACAGAGAGAAGAGGCACTGCGCTTGCGCTATGATTAACAAATGTGTTGTATATTCGTACGC
 TGGCAATAGTTTACAGAACATTTAATACGATATGTATGCTAGCGCGTGACAGGGATGAGTTATATAATAAACT
 CATTGATGCTTTACACTGACTCTGCAGATGATTGCTTTATTGCTAAAGCATATGAATATGCTTTAGCACGC
 CATTGGGACAAAGTTTCAGAAAAATACTTAAGGATGTTTTATGAGAATTGTAGTAAATAA

In specific aspects, the one or more heterologous glycosyltransferase(s) capable of
 10 synthesizing a β -1,2 mannan polymer has an amino acid sequence at least 80%, 90%, 95%, 98%, or
 99% identical to the amino acid sequence of WbaD of *Citrobacter freundii* P079F I comprising SEQ ID
 NO: 11. In other embodiments, the one of more heterologous glycosyltransferase(s) has an amino
 acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of WbaC
 of *Citrobacter freundii* P079F I comprising SEQ ID NO: 12. In still other embodiments, the one of
 15 more heterologous glycosyltransferase(s) has an amino acid sequence at least 80%, 90%, 95%, 98%,
 or 99% identical to the amino acid sequence of WbaB of *Citrobacter freundii* P079F I comprising SEQ
 ID NO: 13.

In other embodiments, the host cell of the invention further comprises a heterologous
 translocase, wherein the heterologous translocase has an amino acid sequence at least 80%, 90%,
 20 95%, 98%, or 99% identical to the amino acid sequence of Wzx of *Citrobacter freundii* P079F I
 comprising SEQ ID NO: 14

In additional embodiments, host cells of the invention comprise one or more polynucleotide
 sequences that encode one or more heterologous guanylyltransferases. In certain embodiments, the
 one or more heterologous guanylyltransferases includes, without limitation, manB and manC. In
 25 certain embodiments, the manB and manC are from *E. coli*. In specific aspects, the *E. coli* is *E. coli*
 K12 W3110. In certain embodiments, the one or more heterologous guanylyltransferases comprises
 manB and manC from *E. coli*, optionally from *E. coli* K12 W3110.

In some aspects, the one or more heterologous guanylyltransferases has an amino acid
 sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of manB of *E.*
 30 *coli* K12 W3110. In some aspects, the one or more heterologous guanylyltransferases has an amino
 acid sequence identical to the amino acid sequence of manB of *E. coli* K12 W3110. In certain
 embodiments, manB of *E. coli* K12 W3110 comprises the amino acid sequence of SEQ ID NO: 31.
 Thus, in certain aspects, the one or more heterologous guanylyltransferases has an amino acid
 sequence at least 80%, 90%, 95%, 98%, or 99% identical to SEQ ID NO: 31. In other aspects, the
 35 one or more heterologous guanylyltransferases has an amino acid sequence identical to SEQ ID NO:
 31:

MKKLTCKFAYDIRGKLGEELNEDIAWRIGRAYGEFLPKPTIVLGGDVRLTSETLKLALAKGLQDAGVDVL
DIGMSGTEEIYFATFHLGVDGGIEVTASHNPM DYNGMKLVREGARPISGDTGLRDVQRLAEANDFPVDET
RYQQINLRDAYVDHLFGYINVKNLTPLKLVINSNGAAGPVVDAIEARFKALGAPVELIKVHNTPDGNFPNGIPNP
LLPECRDDTRNAVIKHGADMGI AFDGDFDRCFLFDEKGQFIEGYYIVGLLAEAFLEKNPGAKIIHDPRLSWNTVD
5 VTAAGGTPVMSKTGHAFIKERMKEDAIYGGEMSAHHYFRDFAYCDSGMIPWLLVAELVCLKDKTLGELVRDR
MAAFPASGEINSKLAQPVEAINRVEQHFSREALAVDRTDGISMTFADWRFNLRSTNTEPVRLNVESRGDVPLM
EARTRTLTLTLLNE

In certain embodiments, host cells of the present invention comprise a nucleotide sequence encoding manB, optionally manB of *E. coli* K12 W3110, optionally a nucleotide sequence encoding
10 manB of *E. coli* K12 W3110 comprising an amino acid sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 31, optionally within a plasmid.

In additional embodiments, host cells of the present invention comprise a nucleotide sequence encoding manB, optionally manB of *E. coli* K12 W3110, optionally a nucleotide sequence encoding
manB of *E. coli* K12 W3110 comprising a nucleotide sequence at least 80%, 85%, 90%, 95%, 96%,
15 97%, 98%, 99% or 100% identical to SEQ ID NO: 33:

ATGAAAAATTAACCTGCTTTAAAGCCTATGATATTCGCGGGAAATTAGGCGAAGAACTGAATGAAG
ATATCGCCTGGCGCATTGGTCGCGCTATGGCGAATTTCTCAAACGAAAACATTGTGTTAGGCGGTGATG
TCCGCCTCACCAGCGAAACCTTAAACTGGCGCTGGCGAAAGTTTACAGGATGCGGGCGTTGACGTGCTGG
ATATTGGTATGTCCGCGACCGAAGAGATCTATTTGCCACGTTCCATCTCGCGTGGATGGCGGCATTGAAG
20 TTACCGCCAGCCATAATCGGATGGATTATAACGGCATGAAGCTGGTTCGCGAGGGGGCTCGCCGATCAGCG
GAGATACCGGACTGCGCGACGTCCAGCGTCTGGCTGAAGCCAACGACTTTCCTCCCGTCGATGAAACCAAAC
GCGGTGCTATCAGCAAATCAACCTGCGTGACGCTTACGTTGATCACCTGTTGCGTTATATCAATGTCAAAAA
CCTCACGCCGCTCAAGCTGGTATCAACTCCGGGAACGGCGAGCGGGTCCGGTGGTGGACGCCATTGAAG
CCCGCTTTAAAGCCCTCGGCGCGCCCGTGAATTAATCAAAGTGACAACACGCCGGACGGCAATTTCCCCA
25 ACGGTATTCCTAACCCTACTGCCGGAATGCCGCGACGACACCCGCAATGCGGTCATCAAACACGGCGCGG
ATATGGGCATTGCTTTTGATGGCGATTTTGACCGCTGTTTCTGTTTGACGAAAAAGGGCAGTTTATTGAGG
GCTACTACATTGTCGGCCTGTTGGCAGAAGCATTCTCGAAAAAATCCCGGCGCGAAGATCATCCACGATC
CACGTCTCTCTGGAACACCGTTGATGTGGTGACTGCCGAGGTGGCACGCCGTAATGTGAAAACCGGAC
ACGCCTTTATTAAGAAGCTATGCGCAAGGAAGACCCATCTATGGTGGCGAAATGAGCGCCCACTTACT
30 TCCGTGATTTGCTTACTGCGACAGCGGCATGATCCCGTGGCTGCTGGTGGCGAACTGGTGTGCTGAAAG
ATAAAACGCTGGGCGAACTGGTACGCGACCGGATGGCGGCGTTTCCGGCAAGCGGTGAGATCAACAGCAAA
CTGGCGCAACCCGTTGAGGCGATTAAACGCGTGGAACAGCATTTTAGCCGTGAGGCGCTGGCGGTGGATCG
CACCGATGGCATCAGCATGACCTTTGCCGACTGGCGCTTTAACCTGCGCACCTCCAATACCGAACCGGTGGT
GCGCCTGAATGTGGAATCGCGCGGTGATGTGCCGTGATGGAAGCGCGAACGCGAACTCTGCTGACGTTGC
35 TGAACGAGTAA

In other aspects, the one or more heterologous guanylyltransferases has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of manC of *E. coli* K12 W3110. In some aspects, the one or more heterologous guanylyltransferases has an amino acid sequence identical to the amino acid sequence of manC of *E. coli* K12 W3110. In certain
 5 embodiments, manC of *E. coli* K12 W3110 comprises the amino acid sequence of SEQ ID NO: 32. Thus, in certain aspects, the one or more heterologous guanylyltransferases has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to SEQ ID NO: 32. In other aspects, the one or more heterologous guanylyltransferases has an amino acid sequence identical to SEQ ID NO: 32:

10 MAQSKLYPVVMAGGSGSRLWPLSRVLYPKQFLCLKGDLTMLQTTCRLNGVECESPVVICNEQHRFIVA
 EQLRQLNKLTEIILEPAGRNTAPAIALAALAAKRHSPESDPLMLVLAADHVIADDAFRAAVRNAMPYAEAGKLV
 TFGIVPDLPETGYGYIRRGESVAGEQDMVAFEVAQFVEKPNLETAQAYVASGEYYWNSGMFLFRAGRYLEELKK
 YRPDILDACEKAMSAVDPLNFIRVDEEAFLACPEESVDYAVMERTADAVVPM DAGWSDVGSWSSLWEISAH
 TAEGNVCHGDVINHKTENSYYAESGLVTTVGKDLVVVQTKDAVLIADRNAVQDVKKVVEQIKADGRHEHRV
 15 HREYRPWGKYDSIDAGDRYQVKRITVKPGEGLSVQMHHHRAEHVWVWAGTAKVTIDGDIKLLGENESIYIPLG
 ATHCLENPGKIPLDLIEVRSGSYLEEDDVRFADRYGRV

In certain embodiments, host cells of the present invention comprise a nucleotide sequence encoding manB, optionally manC of *E. coli* K12 W3110, optionally a nucleotide sequence encoding manC of *E. coli* K12 W3110 comprising an amino acid sequence at least 80%, 85%, 90%, 95%, 96%,
 20 97%, 98%, 99% or 100% identical to SEQ ID NO: 32, optionally within a plasmid.

In additional embodiments, host cells of the present invention comprise a nucleotide sequence encoding manC, optionally manC of *E. coli* K12 W3110, optionally a nucleotide sequence encoding manC of *E. coli* K12 W3110 comprising a nucleotide sequence at least 80%, 85%, 90%, 95%, 96%,
 97%, 98%, 99% or 100% identical to SEQ ID NO: 34:

25 ATGGCGCAGTCGAAACTCTATCCAGTTGTGATGGCAGGTGGCTCCGGTAGCCGCTTATGGCCGCTT
 TCCCGCGTACTTTATCCCAAGCAGTTTTTATGCCTGAAAGGCGATCTCACCATGCTGCAAACCAACCATCTGCC
 GCCTGAACGGCGTGGAGTGCGAAAGCCCGGTGGTGATTTGCAATGAGCAGCACCGCTTTATTGTCGCGGAA
 CAGCTGCGTCAACTGAACAACTTACCGAGAACATTATTCTCGAACCGGCAGGGCGAAACACGGCACCTGCC
 ATTGCGCTGGCGGCGCTGGCGGCAAAACGTCATAGCCCGGAGAGCGACCCGTTAATGCTGGTATTGGCGGC
 30 GGATCATGTGATTGCCGATGAAGACGCGTTCGCTGCCCGCTGCGTAATGCCATGCCATATGCCGAAGCGG
 GCAAGCTGGTGACCTTCGGCATTGTGCCGATCTACCAGAAACCGGTTATGGCTATATTGTCGCGGTGAAG
 TGTCTGCGGGTGAGCAGGATATGGTGGCCTTTGAAGTGGCGCAGTTTGTGAAAAACCGAATCTGGAAACC
 GCTCAGGCCTATGTGGCAAGCGGCGAATATTACTGGAACAGCGGTATGTTCTGTTCCGCGCCGGACGCTAT
 CTCGAAGAACTGAAAAATATCGCCCGGATATCCTCGATGCCTGTGAAAAAGCGATGAGCGCCGTCGATCCG
 35 GATCTCAATTTTATTGCGTGGATGAAGAAGCGTTTCTCGCTGCCCGGAAGAGTCGGTGGATTACGCGGTC
 ATGGAACGTACGGCAGATGCTGTTGTGGTGCCGATGGATGCGGGCTGGAGCGATGTTGGCTCCTGGTCTTC

ATTATGGGAGATCAGCGCCACACCGCCGAGGGCAACGTTTGCCACGGCGATGTGATTAATCACAAAAGTGA
 AAACAGCTATGTGTATGCTGAATCTGGCCTGGTCACCACCGTCGGGGTGAAAGATCTGGTAGTGGTGCAGAC
 CAAAGATGCGGTGCTGATTGCCGACCGTAACGCGGTACAGGATGTGAAAAAGTGGTCGAGCAGATCAAAG
 CCGATGGTCGCCATGAGCATCGGGTGCATCGCGAAGTGTATCGTCCGTGGGGCAAATATGACTCTATCGACG
 5 CGGGCGACCGCTACCAGGTGAAACGCATCACCGTGAAACCGGGCGAGGGCTTGTGCGGTACAGATGCACCAT
 CACCGCGCGGAACACTGGGTGGTTGTCGCGGGAACGGCAAAAGTCACCATTGATGGTGATATCAAAGTCTT
 GGTGAAAACGAGTCCATTTATATTCGCTGGGGGCGACGCATTGCCTGGAAAACCCGGGGAAAATTCCGCTC
 GATTTAATTGAAGTGCCTCCGGCTCTTATCTCGAAGAGGATGATGTGGTGCCTTCGCGGATCGCTACGGA
 CGGGGTGTAA

10

In certain embodiments, the eukaryotic glycosyltransferase capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain is Bmt3. In certain aspects the Bmt3 is from *Candida*. In some aspects, the *Candida* is *C. albicans*. In specific embodiments, the glycosyltransferase capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain is Bmt3 of *C. albicans*. In certain
 15 embodiments, the glycosyltransferase capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of Bmt3 of *C. albicans*. In other aspects, the glycosyltransferase capable of covalently bonding a mannose molecule to a β -1,2 mannan
 20 polymer to extend a β -1,2 mannan polymer chain has an amino acid sequence identical to the amino acid sequence Bmt3 of *C. albicans*. In certain embodiments, periplasmic expression of Bmt3 of *C. albicans* in the host cell is required to extend a β -1,2 mannan polymer chain. Thus, in certain aspects, addition of the terminal mannose residue to the β -1,2 mannan polymer chain requires periplasmic expression of Bmt3 of *C. albicans* in the host cell. In other embodiments, addition of the terminal
 25 mannose residue to the β -1,2 mannan polymer chain further requires presence of GDP-mannose in the culture medium. In certain aspects, the GDP-mannose is added to the harvested culture. Thus, in certain aspects, the conditions suitable for the production of a glycoconjugate of the invention includes the addition of GDP-mannose to the culture medium. In specific aspects, the GDP-mannose is added to the harvested culture.

30

In certain embodiments, the Bmt3 of *C. albicans* comprises the amino acid sequence of SEQ ID NO: 15. In other embodiments, the Bmt3 of *C. albicans* comprises an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to SEQ ID NO: 15:

MKVKVLSLLVPALLVAGAANASDYTPIKVSGYTFKNQVATKNLQCDSIVYDQDLQVSQAVDLNKPEDLKFFRD
 KLNELRSLNNIYDLFFQDNEDEVEESILERKWKYKCGSAVWLDKYGVYFMVNRIAYSCKGTRNNPTISVLAGQVF
 35 DKNWIELTGKKFPFSGLEFPPTILPHYIDEGKEAEKVLGAEDPRVILHEYTNENGIRIQEPLIAFNALSTEVWDWKRA
 MHIYRPLHDPHRIIRLSIENYAPREKEKNWAPFIDGNLNFVYNFPLRILRCNINNGDCQKVSQKVPDFNDKSHENA

GKLRGGTNLVEIPSQSLPKHLRSRKYWFGIARSHITDCGCVGELYRPHLILISRNKKSQYELNYSVDLIDFNVNP
EPWTPGKTTCSDGKSVLIPNSVAFIKDDYMSVTFSEADKTNKLINAKGWLTITKMLEFTQERLKDESSDPVLES
RLLSKCSTFLAQQYCALSKDTMGWDKLSR

In certain embodiments, host cells of the present invention comprise a nucleotide sequence
5 encoding Bmt3, optionally Bmt3 of *C. albicans*, optionally a nucleotide sequence encoding Bmt3 of *C.
albicans* comprising an amino acid sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%
or 100% identical to SEQ ID NO: 15, optionally within a plasmid.

In other embodiments, host cells of the present invention comprise a nucleotide sequence
encoding Bmt3, optionally Bmt3 of *C. albicans*, optionally a nucleotide sequence encoding Bmt3 of *C.
10 albicans* comprising a nucleotide sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or
100% identical to SEQ ID NO: 29:

ATGAAAGTGAAAGTGCTGAGCCTGCTGGTCCGGCACTGCTGGTTGCAGGTGCCGCCAATGCTAGCGACTAC
ACCCGATTAAAGTGAGCGTTATACCTTCAAAAACAGGTTGCGACCAAGAACCTGCAATGCGATAGCATC
GTGTACGACCAGGATCTGGACCTGCAGGTGTCTCAAGCGTTGATCTGAACAAGCCGAGGACCTGAAATTC
15 TTTCGTGATAAGCTGAACGAACTGCGTAGCCTGAACAACATTTACGACCTGTTCTTTCAAGATAACGAGGAC
GAAGTGGAGGAAAGCATCCTGGAGCGTAAGTGGTATAAATTCTGCGGTAGCGCGGTTTGGCTGGATAAATA
CGGCGTGATTTTATGGTTAACCGTATCGCGTATAGCAAGAAAGGTACCCGTAACAACCCGACCATTAGCGT
GCTGGCGGGCCAGGTTTTTCGACAAGAACTGGATCGAGCTGACCGGTAAGAAATCCCGTTTAGCGGCCTGG
AGTTTTCCGACCATCTGCCGCACTACATTGACGAGGGTAAAGAGGCGGAAAAAGTGATCCTGGGCGCGGAA
20 GATCCGCGTGTTATTCTGCACGAGTATACCAACGAAAACGGTATCCGTATTCAAGAGCCGCTGATTGCGTTC
AACGCGCTGAGCACCGAAGTTGACTGGAAACGTGCGATGCACATTTACCGTCCGCTGCACGATCCGCACCGT
ATCATTCGTCTGAGCATCGAAAATATGCGCCGCGTGAGAAGGAGAAGAACTGGGCGCCGTTTCATCGACGGC
AACAACTGAACTTCGTGTACAACTTTCCGCTGCGTATCCTGCGTTGCAACATTAAACAGGTGATTGCCAGA
AAGTTAGCGGCCCGGATTTTAACGACAAAAGCCACGAGAACGCGGGCAAGCTGCGTGGTGGCACCAACCTG
25 GTGGAAATCCCGAGCCAAAGCCTGCCGAAACACCTGCGTAGCCGTAAGTACTGGTTCGGCATCGCGCTAGC
CACATTACCGACTGCGGTTGCGTTGGCGAGCTGTATCGTCCGCACCTGATCCTGATTAGCCGTAACAAGAAA
AGCGATCAGTACGAGCTGAACTATGTGAGCGATCTGATCGACTTTAACGTTAACCCGGAACCGTGGACCCCG
GGTAAACACCTGCAGCGACGGCAAGAGCGTGCTGATCCCGAACAGCGTTGCGTTCATTAAGGACGATTAC
ATGAGCGTGACCTTTAGCGAAGCGGATAAGACCAACAACTGATCAACGCGAAAGGTTGGCTGACCTATATT
30 ACCAAGATGCTGGAGTTCACCAAGAAGCTCTGAAAGATGAGAGCAGCGACCCGTTCTGGAAAGCCGCTCTG
CTGAGCAAGTGCAGCACCTTTCTGGCGCAGCAATACTGCGCGCTGAGCAAAGATACCATGGGCTGGGACAAG
CTGAGCCGTAA

Oligosaccharyl Transferase

35 N-linked protein glycosylation (the addition of carbohydrate molecules to an asparagine
residue in the polypeptide chain of the target protein) is the most common type of post-translational

modification occurring in the endoplasmic reticulum of eukaryotic organisms. The process is accomplished by the enzymatic oligosaccharyl transferase complex (OST) responsible for the transfer of a preassembled oligosaccharide from a lipid carrier (dolichol phosphate) to an asparagine residue of a nascent protein within the conserved sequence Asn-X-Ser/Thr (where X is any amino acid except proline) in the Endoplasmic Reticulum.

It has been shown that a bacterium, the food-borne pathogen *Campylobacter jejuni*, can also N-glycosylate its proteins (Wacker *et al.* Science. 2002; 298(5599):1790-3) due to the fact that it possesses its own glycosylation machinery. The machinery responsible of this reaction is encoded by a cluster called "pgl" (for protein glycosylation). In certain embodiments, the *C. jejuni* glycosylation machinery is transferred to *E. coli* to allow for the glycosylation of recombinant proteins expressed by the host *E. coli* cells. Previous studies have demonstrated how to generate *E. coli* strains that can perform N-glycosylation (see, *e.g.* Wacker *et al.* Science. 2002; 298 (5599):1790-3; Nita-Lazar *et al.* Glycobiology. 2005; 15(4):361-7; Feldman *et al.* Proc Natl Acad Sci U S A. 2005; 102(8):3016-21; Kowarik *et al.* EMBO J. 2006; 25(9):1957-66; Wacker *et al.* Proc Natl Acad Sci U S A. 2006; 103(18):7088-93; International Patent Application Publication Nos. WO2003/074687, WO2006/119987, WO 2009/104074, and WO/2011/06261, and WO2011/138361).

Thus, in certain aspects, the host cells of the present invention comprise a nucleotide sequence encoding a heterologous oligosaccharyl transferase, optionally within a plasmid. In some embodiments, the heterologous oligosaccharyl transferase is a PglB. In certain embodiments, the PglB is from *Campylobacter*. In certain aspects, the *Campylobacter* includes, but is not limited to, *Campylobacter jejuni* or *Campylobacter coli*. In some aspects, the *Campylobacter* is *Sinorhizobium meliloti* 1021. In certain embodiments, the PglB is optionally from *Campylobacter*, optionally from *Campylobacter jejuni* or *Campylobacter coli*, optionally an evolved PglB comprising an amino acid sequence of SEQ ID NO: 16. In certain embodiments, the PglB is an evolved pglB comprising an amino acid sequence at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of PglB of *Campylobacter coli*. In other aspects, the PglB comprises an amino acid sequence identical to the amino acid sequence of PglB of *Campylobacter coli*. In certain embodiments, the PglB is an evolved PglB enzyme comprising the amino acid sequence of SEQ ID NO: 16. In certain aspects, the evolved PglB has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to SEQ ID NO: 16. In other aspects, the evolved PglB has an amino acid sequence identical to SEQ ID NO: 16:

MLKKEYLKNPYLVLFAMIIAYVFSVFCRFYWWWASEWNEYFHNNQLMIISNDGYAFAEGARDMIAG
FHQPNDLSYYGSSLSALTYWLYKITPFSFESIILYMSTFLSSLVIPQILLANEYKRRLMGFVAALLASIANSYNRT
MSGYYDTDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFIGIYLWWYPSSYTLNVALIGLFLIYTLIFHRKEKIFYIA
VILSSLTSLNIAWFYQSAIIVILFALFALEQKRLNFMIIIGILGSAWLIFLILSGGVDPILYPLKFYIFRSDESTNLTQGF
MYFNVVQTIQEVENVDLSEFMRRISGSEIVFLFSLLGFWLLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMA

LGFGFLLSEFKAIMVKKYSQLTSCVCIVFATILTLAPVFIHIYNYRAPTVFSQNEASLLNQLKNIANREDYVVTWW
 DYGYPVRYSDVKTMVDGGKHLGKDNFFPSFALSQDEQAANMARLSVEYTEKSFYAPQNDILKTDILQAMMK
 DYNQSNVDLFLASLSKPDFKIDTPKTRDIYLYMPARMSLIFSTVANFSFINLDTGVLDKPFTFSTAYPLDVKNGEIY
 LSNQVVLSDDFRSFKIGDNNVSVNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFNSAYVQ
 5 MFFLGNYDKNLFDLVINSRDAKVFLKI

In certain embodiments, host cells of the present invention comprise a nucleotide sequence encoding an evolved PglB, optionally a nucleotide sequence encoding an evolved PglB comprising an amino acid sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 16, optionally within a plasmid.

10 In other embodiments, host cells of the present invention comprise a nucleotide sequence encoding an evolved PglB, optionally a nucleotide sequence encoding an evolved PglB comprising a nucleotide sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 30:

ATGCTGAAGAAGGAATATCTGAAGAACCCGTATCTGGTGCTGTTTGCGATGATTATCCTGGCGTAT
 15 GTTTTTAGTGTGTTTTGTCGTTTCTACTGGGTGTGGTGGGCCAGTGAATGGAACGAATATTTCCACAACAAC
 CAGCTGATGATCATCTCCAATGATGGCTATGCCTTCGCAGAAGGTGCCCGTGACATGATTGCAGGCTTTCAT
 CAGCCGAACGATCTGAGTTATTACGGTAGCTCTCTGTCCGCGCTGACCTATTGGCTGTACAAAATCACGCCG
 TTTAGTTTCGAATCCATTATCCTGTACATGAGTACCTTCCTGAGTTCCTGGTGGTTATTCCGAGATCCTGC
 TGGCCAATGAATATAAACGTCGGCTGATGGGCTTTGTTGCGGCCCTGCTGGCTAGTATTGCGAACTCCTATT
 20 ACAATCGCACCATGAGTGGTTATTACGATACGGACATGCTGGTCATTGTGCTGCCGATGTTTCATCCTGTTTT
 TCATGGTGCGTATGATTCTGAAAAAGGATTTCTTTAGCCTGATCGCCCTGCCGCTGTTTATTGGCATCTATCT
 GTGGTGGTACCCGTCATCGTATACCCTGAACGTTGCACTGATTGGTCTGTTTCTGATTTACACGCTGATCTTC
 CATCGCAAGGAAAAGATCTTTTATATCGCGTTATCTTAAGCTCTCTGACCCTGAGCAACATTGCTTGGTTTT
 ATCAGTCTGCGATTATCGTCATCCTGTTTGCCCTGTTGCACTGGAACAAAAACGTCTGAATTTTCATGATTAT
 25 CGGCATTCTGGGTAGTGCCTGGCTGATCTTTCTGATTCTGTCCGGCGGTGTTGATCCGATTCTGTACCCGCT
 GAAATTTTATATCTTCCGCTCAGATGAGTCGACCAACCTGACCCAAGGCTTCATGTACTTCAACGTTGTGCAG
 ACGATCCAAGAAGTGGAATGTTGATCTGAGCGAATTTATGCGTCGCATTAGTGGCTCCGAAATCGTTTTT
 CTGTTCTCACTGTTAGGTTTCGTCTGGCTGCTGCGTAAACACAAGTCGATGATTATGGCCCTGCCGATTCTG
 GTGCTGGGTTTCTGGCACTGAAAGGCGGTCTCCGCTTTACCATTTACAGCGTTCGGTTCATGGCCCTGGGC
 30 TTTGGTTTCTGCTGTCTGAATTTAAGGCAATCATGGTTAAAAAGTACTCACAGCTGACCTCGTGGTCTGCA
 TTGTGTTGCCACCATCCTGACGCTGGCACCAGGTGTTTCATCCATATCTACAACTACAGGGCTCCGACGGTGT
 TTAGCCAGAACGAAGCGTCGCTGCTGAATCAACTGAAGAACATTGCCAATCGTGAAGATTATGTCGTGACCT
 GGTGGGACTATGGCTACCCGGTGGCTATTACAGCGATGTTAAACGATGGTCGACGGCGGTAACACCTG
 GGCAAGGATAACTTTTTCCCGAGCTTTGCTCTGTCTAAAGATGAACAGGCAGCTGCGAATATGGCGCGCCTG
 35 TCAGTCGAATACCCGAAAAGTCGTTTTATGCCCGCAGAATGATATTCTGAAAACGGACATCCTGCAAGCA
 ATGATGAAGGATTATAACCAAAGCAATGTTGACCTGTTCTGGCATCACTGTGAAACCGGATTTTAAAGATTG

ACACCCCGAAAACGCGTGATATCTATCTGTACATGCCGGCTCGCATGAGTCTGATTTTTAGCACCGTCGCGA
 ACTTTTCTTTTCATCAACCTGGATACGGGCGTGCTGGACAAACCGTTTACCTTCTCAACGGCGTACCCGCTGG
 ATGTGAAGAACGGCGAAATTTATCTGTCAATGGTGTTCCTGAGCGATGACTTTCGTTCTTTCAAATCG
 GCGATAACGTTGTGAGCGTGAACAGCATCGTTGAAATTAATAGCATCAAACAGGGTGAATACAAGATTACCC
 5 CGATCGATGACAAGGCTCAATTCTACATTTTCTACCTGAAGGACTCCGCTATTCCGTATGCGCAGTTCATCCT
 GATGGATAAAACCATGTTTAACTCTGCGTACGTGCAAATGTTTTTCTGGGTAACACGATAAGAACCTGTTT
 GACCTGGTCATTAATTCTCGCGATGCTAAGGTGTTTAACTGAAGATCTAA

In certain aspects, the nucleotide sequence encoding PglB is codon optimized.

Polymerase

10 Host cells of the present invention may also comprise a nucleotide sequence that encodes a polymerase (*e.g. wzy*). In certain embodiments, the polymerase (*e.g. wzy*) is introduced into a host cell of the invention (*i.e.* the polymerase is heterologous to the host cell). In some embodiments, the polymerase is a bacterial polymerase. In other embodiments, the polymerase is a capsular polysaccharide polymerase (*e.g. wzy*) or an O antigen polymerase (*e.g. wzy*). In certain aspects, the
 15 polymerase is an O-antigen polysaccharide polymerase (*e.g. wzy*), *e.g.* from *Shigella* species, *Pseudomonas* species or *Escherichia* species. (*e.g. Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, or *E. coli*). In other aspects, the polymerase is a capsular polysaccharide polymerase (*e.g. wzy*), *e.g.* from *N. meningitidis* serogroup A (MenA), *N. meningitidis* serogroup C (MenC), *N. meningitidis* serogroup Y (MenY), *N. meningitidis* serogroup W (MenW), *H. influenzae* type b (Hib), Group B *Streptococcus* (GBS), *Streptococcus pneumoniae*, or *Staphylococcus aureus*. In yet ther apsects, the polymerase is a capsular polysaccharide polymerase (*e.g. wzy*) of
 20 *Streptococcus pneumoniae*. In some mebodiments, said *wzy* polymerase may be incorporated (*e.g.* inserted into the genome or expressed by a plasmid) in said host cell as part of a *rfb* cluster or capsular polysaccharide cluster. Thus, in certain embodiments, a host cell of the invention may further
 25 comprise a nucleotide sequence encoding a heterologous *wzy* polymerase.

Flippases

A host cell of the invention may also comprise a nucleotide sequence encoding a heterologous flippase (*e.g. wzx*), *e.g.* a heterologous flippase. Flippases translocate wild type repeating units and/or
 30 their corresponding engineered (hybrid) repeat units from the cytoplasm into the periplam of host cells (*e.g. E. coli*). In certain embodiments, the flippase is a bacterial flippase, *e.g.* a flippase of the polysaccharide biosynthetic pathway of interest. In some embodiments, the host cell of the invention comprises a nucleotide sequence encoding a flippase (*e.g. wxz* gene) of a polysaccharide biosynthetic pathway of a *Streptococcus* species, *Shigella* species, *Escherichia* species, *Pseudomonas* species, or
 35 *Staphylococcus* species. (*e.g. Streptococcus pneumoniae*, *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *E. coli*, *Pseudomonas aeruginosa*, or *Staphylcoccus aureus*. In other embodiments,

the flippase is a capsular polysaccharide flippase (*e.g. wzx*) of *Streptococcus pneumoniae*. Other flippases that can be introduced into the host cells of the invention are for example from *Campylobacter jejuni* (*e.g. pglk*) or Wzx of *Citrobacter* (*e.g., Citrobacter freundii* P079F I).

Translocases

5 A host cell of the invention may also comprise a nucleotide sequence encoding a translocase (*e.g. Wzx*), *e.g.* a heterologous translocase. In certain aspects, the translocase is a Wzx/Wzy dependent transporter. In other aspects, the translocase is a ATP-binding cassette (ABC) dependent transporter. In yet other aspects, the translocase is a synthase dependent transporter. In specific
10 embodiments, the translocase is an ABC transporter. Polysaccharides assembled by ABC transporters are fully polymerized by sequential glycosyl transfer at the cytoplasmic face of the inner membrane of a host cell. The glycan can be assembled as an Undecaprenyl diphosphate (Und-PP)-linked intermediate (*e.g.* for most O-antigen polysaccharides). Alternatively, for certain types of
15 polysaccharides (*e.g.* capsular polysaccharides), the acceptor is diacylglycerol phosphate. The polysaccharide chain is elongated by addition of monomers to the nonreducing terminus of the lipid-linked intermediate and is completed in the cytoplasm, prior to export to the periplasmic space via the
20 translocase (*e.g.* ABC transporter) (Cuthbertson L. et al., 2010, *Microbiology and Molecular Biology Reviews*, 74(3):341-362; Bi Y. et al., 2018, *Nature*, 553(7688):361-365)).

 In certain embodiments, a saccharide antigen of the invention is synthesized by a ABC transporter-dependent pathway. Thus, in certain aspects, a saccharide antigen of the invention is
25 completely synthesized on the cytosolic leaflet of the plasma membrane of a host cell (*e.g. E. coli*). In certain embodiments, the saccharide antigen of the invention is built on a lipid acceptor (*e.g.* Undecaprenyl diphosphate (Und-PP)) leading to the formation of a lipid-linked antigen saccharide. In certain aspects, Und-PP serves as a lipid acceptor and is modified by the addition of an acetylated amino sugar phosphate (*e.g., N-acetylglucosamine-1-P*) to generate a biosynthesis primer. In specific
30 embodiments, a translocase (*e.g. Wzx*) translocates the Und-PP-linked saccharide antigen intermediate to the membrane's periplasmic side, where it forms a substrate for glycosylation of a carrier protein (*e.g.* a modified Sap2 protein of the invention). In certain embodiments, the translocase is a bacterial translocase, *e.g.* a translocase of the polysaccharide biosynthetic pathway of interest. Thus, in some embodiments, the host cell of the invention comprises a nucleotide
35 sequence encoding a translocase (*e.g. wzm-wzt*) of a polysaccharide biosynthetic pathway of a *Klebsiella* species, *Streptococcus* species, *Shigella* species, *Escherichia* species, *Pseudomonas* species, or *Staphylococcus* species. (*e.g. Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *E. coli*, *Pseudomonas aeruginosa*, or *Staphylococcus aureus*). In specific embodiments, the heterologous translocase that is introduced into a host cell of
the invention is an ABC transporter (*e.g. wzm-wzt*) of *Klebsiella pneumoniae*. Other translocases that

can be introduced into the host cells of the invention are for example from *Campylobacter jejuni* (e.g. *pglK*) or Wzx of *Citrobacter* (e.g., *Citrobacter freundii* P079F I).

In some embodiments, the heterologous translocase includes, without limitation, Wzx from *Citrobacter*. In specific aspects, the *Citrobacter* is *Citrobacter freundii*. In certain aspects, the *Citrobacter freundii* is *Citrobacter freundii* P079F I. In certain embodiments, the translocase is Wzx from *Citrobacter*, optionally from *Citrobacter freundii*, optionally *Citrobacter freundii* P079F. In certain aspects, the heterologous translocase has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of Wzx of *Citrobacter freundii* P079F I. In some aspects, the heterologous translocase has an amino acid sequence identical to the amino acid sequence of Wzx of *Citrobacter freundii* P079F I. In certain embodiments, Wzx of *Citrobacter freundii* P079F I comprises the amino acid sequence of SEQ ID NO: 14. Thus, in certain aspects, the heterologous translocase has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to SEQ ID NO: 14. In other aspects, the heterologous translocase has an amino acid sequence identical to SEQ ID NO: 14:

MCTKFIKKIPSHFVAGSAWGSRFISIFVQFYKILLNYLGTNGYALFSLIASFSAWFLLDIGMSTNLQ
NKISERKAYGKAYFDLVKRTGCFILVALTLFVILLWIFGPYLSRILLVSFDLSEKDKNNIFFISSLLFIGNGIGFFAYK
IWYAEHKGWISNILPAISSICGLLFLILLRTEQFNISHLIIVCLLSFYGPAAFLGVLSTFTFASSLHYKEKPLSHSIVQ
DIIRPSMGFFSFSVMAALVLQVDYIIMSHTLTGKDIVINYVLSKIFGLINFLYAALLQSVWPLCAEAKYKDDKSIYN
QIKVKYIGFGAILVFAISLFIYLARDLIILLARDAGISFSFTLIAFFAIYHMIRIWTDTYAMFLLSTGNMRFLYISVPC
QALLSGVLQWYGSVLFGLPGIILGLIMSYILTVSWILPYSFNKNVQGRLCR

In certain embodiments, host cells of the present invention comprise a nucleotide sequence encoding Wzx, optionally Wzx of *Citrobacter freundii* P079F I, optionally a nucleotide sequence encoding Wzx of *Citrobacter freundii* P079F I comprising an amino acid sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 14, optionally within a plasmid.

In additional embodiments, host cells of the present invention comprise a nucleotide sequence encoding Wzx, optionally Wzx of *Citrobacter freundii* P079F I, optionally a nucleotide sequence encoding Wzx of *Citrobacter freundii* P079F I comprising a nucleotide sequence at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 28:

ATGTGCACTAAATTTATAAAAAAATCCCTAGTCATTTTGTGTAGCAGGTAGTGCCTGGGGGAGTC
GATTTATATCTATATTTGTACAATTCTATAGTATAAAGATATTATTAACATTTTGGGTACAAACGGGTATGC
TCTTTTCTCATTAATTGCGAGTTTCTCTGCATGGTTTCTCTTGTTGATATTGGTATGTCAACCAATTTGCAA
AATAAAATATCGGAGCGTAAAGCCTATGGAAAGGCATATTTGATTTGGTTAAAGAACTGGTTGTTTTTTAA
TTGTAGCTTTGACTCTATTTGTAATTTGCTTTGGATATTTGGTCCTTATTTATCGAGGATATTATTGGTGTC
GTTTGATTTTTTATCTGAAAAAGACAAAAATAATATTTCTTTATTTTCATCTCTGTTATTCATAGGTAACGGCA
TTGGTTTTTTTGCATACAAAATTTGGTATGCAGAACATAAAGTTGGATTCTAATACTACTACCCGCTATATC
TTCCATATGTGGGTGTTGTTTTTAATCTTACTAAGAACTGAACAATTTAATAAGTCATTTAATTATTGTAT
GCCTCTTATCATTCTATGGTCCCGCAGCTTTTCTGGGTGTAAGTTTATTTTAAACAACATTTGCTAGTTCGCT

ACATTATAAAGAAAAACCTTTATCTCACTCCATCGTGCAGGATATTATTAGGCCTTCTATGGGTTTTTTTTTCA
 TTTTCTGTAATGGCTGCATTGGTACTACAAGTAGATTATATTATTATGTCCCATACACTAACAGGAAAAGACA
 TTGTAATATATAATGTGTTAAGTAAAATATTCGGGTAAATCAATTTCTATATGCAGCGTTGCTACAATCAGT
 TTGGCCATTATGTGCAGAAGCAAAATATAAAGATGATAAAAGTATTTACAATCAGATAAAAGTAAAGTATATC
 5 GGCTTTGGTGCAATATTGGTATTCGCAATTTCTTTATTTATCTATCTTGCTAGGGATTTGATTATATTATTAT
 TAGCACGTGATGCAGGTATTTCAATTTCTTTCACTCTCATTGCTTTTTTTGCAATCTATCACATGATAAGAAT
 ATGGACAGATACGTATGCTATGTTTTTACTTAGTACGGGGAATATGAGGTTCTGTATATATCTGTGCCATG
 CCAAGCTTTGTTGAGTGGAGTGTGCAATGGTATGGTTCTGTTTTATTTGGACTTCCAGGGATTATATTAGG
 CTTAATTATGTCTTATATATTAACAGTCAGCTGGATATTACCATATTCATTCAATAAAAAATGTTCAAGGTAGA
 10 TTGTGTCGATAA

Thus, in certain embodiments, the present invention provides a host cell comprising a nucleotide sequence comprising (i) a *wzx* gene comprising a nucleotide sequence of SEQ ID NO: 28, optionally comprising a nucleotide sequence at least 80%, 90%, 95%, 98% or 99% identical to SEQ ID NO: 28. In certain aspects, the *wzx* gene is from *Citrobacter freundii* P079F I.

15 In additional embodiments, the present invention provides a method of producing a β -1,2 mannan polymer in a host cell of the invention, the method comprising the steps of introducing and expressing in the host cell:

- i. a nucleotide sequence encoding one or more first heterologous glycosyltransferase(s) capable of synthesizing a β -1,2 mannan polymer;
- 20 ii. a nucleotide sequence encoding a second heterologous glycosyltransferase which is eukaryotic and is capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain, wherein the second heterologous glycosyltransferase is Bmt3, optionally of *C. albicans*; and
- iii. optionally, a nucleotide sequence encoding a translocase capable of translocating the
 25 β -1, 2 mannan to periplasmic side of an inner membrane of the host cell.

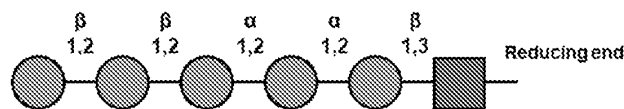
In specific aspects, the present invention provides a method of producing a β -1,2 mannan polymer in a prokaryotic host cell, the method comprising the steps of introducing and expressing in the host cell:

- 30 iv. a nucleotide sequence encoding one or more first heterologous glycosyltransferase(s) capable of synthesizing a β -1,2 mannan polymer, wherein the one of more heterologous glycosyltransferase(s) comprises WbaD, WbaC, and WbaB from *Citrobacter*, optionally from *Citrobacter freundii*, optionally from *Citrobacter freundii* P079F I;

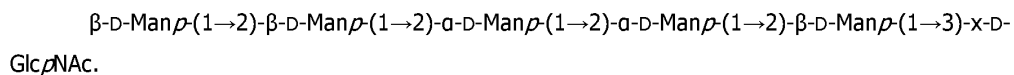
- v. a nucleotide sequence encoding a second heterologous glycosyltransferase which is eukaryotic and is capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain, wherein the second heterologous glycosyltransferase is Bmt3, optionally of *C. albicans*; and
- 5 vi. optionally, a nucleotide sequence encoding a translocase capable of translocating the β -1,3 glucan to periplasmic side of an inner membrane of the prokaryotic host cell, wherein the translocase comprises Wzx from *Citrobacter*, optionally from *Citrobacter freundii* P079F 1,

wherein the β -1,2 mannan polymer comprises at least five consecutive β -1,2 linked mannose
10 molecules.

In certain aspects, the β -1,2 mannan polymer is a fungal β -1,2 mannan polymer. In some aspects, the fungal β -1,2 mannan polymer is from *Candida*. In specific aspects, the fungal β -1,2 mannan polymer is from *Candida albicans*. In certain embodiments, the β -1,2 mannan polymer has the structure:



15 In additional embodiments, the β -1,2 mannan polymer has the structure:



20 In some aspects, the prokaryotic host cell is a bacterial cell. In some aspects, the bacteria is a gram positive bacteria. In other aspects, the bacteria is a gram negative bacteria. In certain embodiments, the bacteria includes, but is not limited to, an *Escherichia* species, a *Shigella* species, *Klebsiella* species, a *Xanthomonas* species, a *Salmonella* species, a *Yersinia* species, a *Lactococcus* species, a *Lactobacillus* species, a *Pseudomonas* species, a *Corynebacterium* species, a *Streptomyces* species, a *Streptococcus* species, a *Staphylococcus* species, a *Bacillus* species, and a *Clostridium* species. In some aspects, the bacteria is an *E. coli* species. In specific aspects, the bacteria is *E. coli*. In certain embodiments, the prokaryotic host cell includes, but is not limited to, an *Escherichia* species, a *Shigella* species, *Klebsiella* species, a *Xanthomonas* species, a *Salmonella* species, a *Yersinia* species, a *Lactococcus* species, a *Lactobacillus* species, a *Pseudomonas* species, a *Corynebacterium* species, a *Streptomyces* species, a *Streptococcus* species, a *Staphylococcus* species, a *Bacillus* species, and a *Clostridium* species. In some aspects, the prokaryotic host cell is an *E. coli* species. In specific
30 embodiments, the prokaryotic host cell is *E. coli*. In certain embodiments, the GlcNAc is linked to a

lipid carrier by WecA from the host *E. coli* cell. In specific embodiments, the lipid carrier is undecaprenyl.

Bioconjugates

In certain embodiments, the present invention provides a bioconjugate comprising a modified Sap2 protein of the invention linked to an antigen, as described herein. In certain aspects, the antigen is linked to an amino acid on the modified Sap2 protein selected from asparagine, aspartic acid, glutamic acid, lysine, cysteine, tyrosine, histidine, arginine or tryptophan (*e.g.* asparagine). Bioconjugates, as described herein, have advantageous properties over chemical conjugates of antigen-carrier protein, in that they require less chemicals in manufacture and are more consistent in terms of the final product generated.

In certain embodiments, the present invention provides a method of producing a bioconjugate that comprises a modified Sap2 protein of the invention linked to at least one saccharide antigen, the method comprising: (1) culturing a host cell of the invention under conditions suitable for the production of proteins; and (2) isolating the bioconjugate produced by a host cell of the invention. In some embodiments, the bioconjugate is isolated or purified from a whole cell extract of the host cell. In other embodiments, the bioconjugate is isolated or purified from a cytoplasmic extract from the host cell. In specific embodiments, the bioconjugate is isolated or purified from a periplasmic extract from the host cell.

Methods of making Bioconjugates are known in the art. For example, Bioconjugates of the invention can be made using the shakeflask process, *e.g.* in a LB shake flask. In certain aspects, a fed-batch process for the production of recombinant glycosylated proteins in bacteria can be used to produce bioconjugates of the invention. The aim is to increase glycosylation efficiency and recombinant protein yield per cell and while maintaining simplicity and reproducibility in the process. Bioconjugates of the invention can be manufactured on a commercial scale by developing an optimized manufacturing method using typical *E. coli* production processes. Various types of feed strategies, such as batch, chemostat and fed-batch can be used.

Thus, in specific embodiments, the present invention provides a bioconjugate produced by a method of the invention. In specific aspects, a bioconjugate of the invention comprises at least one saccharide antigen linked to a modified Sap2 protein of the invention.

The bioconjugates of the invention can be purified for example, by chromatography (*e.g.* ion exchange, anionic exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. See, *e.g.* Saraswat *et al.* 2013, Biomed. Res. Int. ID#312709 (p. 1-18); see also the methods described in WO 2009/104074. Further, the bioconjugates of the invention may be fused to heterologous polypeptide sequences described herein or otherwise known in the art to facilitate purification.

In additional embodiments, the present invention provides a method of producing a glycoconjugate comprising a modified carrier protein and at least one saccharide antigen. As used herein, the term "glycoconjugate" refers to a hybrid molecule composed of a carrier protein and multiple polysaccharide chains, wherein the polysaccharides are covalently linked to the carrier protein.

Thus, in certain embodiments, the present invention provides a method of producing a glycoconjugate comprising a modified carrier protein and a β -1,2 mannan, wherein said method comprises the steps of i) culturing the host cell of the invention under conditions suitable for the production of proteins, ii) harvesting the culture to produce a harvested culture, and iii) isolating the glycoconjugate from the culture. In certain aspects, the conditions suitable for the production of the glycoconjugate includes the addition of GDP-mannose to the culture medium. In specific aspects, the GDP-mannose is added to the harvested culture.

In certain aspects, the modified carrier protein of the invention includes, but is not limited to, Sap2, Als3, detoxified Exotoxin A of *P. aeruginosa* (EPA), CRM197, Diphtheria toxoid, tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A of *S. aureus*, clumping factor B of *S. aureus*, *E. coli* FimH, *E. coli* FimHC, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* sat protein, the passenger domain of *E. coli* sat protein, *C. jejuni* AcrA, and a *C. jejuni* natural glycoprotein. In specific aspects, the modified carrier protein is a modified Sap2 protein of the invention.

In certain embodiments, the at least one saccharide antigen comprises a β -1,2 mannan polymer. In some aspects, the β -1,2 mannan polymer comprises at least two β -1,2 linked mannose molecules. In additional aspects, the β -1,2 mannan polymer comprises two to fifty, two to forty, two to thirty, two to twenty, two to ten, two to nine, two to eight, two to seven, two to six, two to five, two to four, two to three, two, three to fifty, three to forty, three to thirty, three to twenty, three to ten, three to nine, three to eight, three to seven, three to six, three to five, three to four, three, four to fifty, four to forty, four to thirty, four to twenty, four to ten, four to nine, four to eight, four to seven, four to six, four to five, four, five to fifty, five to forty, five to thirty, five to twenty, five to ten, five to nine, five to eight, five to seven, five to six, or five, six to forty, six to thirty, six to twenty, six to ten, six to nine, six to eight, six to seven, six, seven to fifty, seven to forty, seven to thirty, seven to twenty, seven to ten, seven to nine, seven to eight, or seven 1,2 linked mannose molecules. In some aspects, the β -1,2 mannan polymer comprises at least 2, at least 3, at least 4, or at least 5 consecutive β -1,2 linked mannose molecules.

In certain embodiments, the present invention provides a method of producing a glycoconjugate comprising a modified carrier protein and a β -1,2 mannan. In some embodiments,

the method of producing a glycoconjugate comprising a modified carrier protein and a β -1,2 mannan comprises culturing the host cell the invention under conditions suitable for the production of proteins.

In further embodiments, the present invention provides a method of producing a bioconjugate in a host cell of the invention, the method comprising the steps of:

- 5 a. obtaining a host cell of the invention that produces a β -1,2 mannan polymer; and
- b. further introducing and expressing in the host cell:
 - 10 i. a nucleotide sequence encoding a modified carrier protein comprising at least one glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are any amino acid except proline, and wherein the modified carrier protein further comprises an N-terminal bacterial signal sequence capable of transporting the modified carrier protein to the periplasmic side of the inner membrane of the host cell; and
 - 15 ii. a nucleotide sequence encoding an oligosaccharyl transferase capable of producing a bioconjugate by transferring the β -1,2 mannan polymer from a lipid carrier to the modified carrier protein.

In certain aspects, the present invention provides a method of producing a bioconjugate in a prokaryotic host cell, the method comprising the steps of:

- a. obtaining a prokaryotic host cell of the invention that produces a β -1,2 mannan polymer; and
- 20 b. further introducing and expressing in the host cell:
 - 25 i. a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are any amino acid except proline, and wherein the modified carrier protein further comprises an N-terminal bacterial signal sequence capable of transporting the modified carrier protein to the periplasmic side of the inner membrane of the prokaryotic host cell; and
 - 30 ii. a nucleotide sequence encoding an oligosaccharyl transferase capable of producing a bioconjugate by transferring the β -1,2 mannan polymer from a lipid carrier to the modified carrier protein, wherein the oligosaccharyl transferase is PglB from *Campylobacter*, optionally from *Campylobacter jejuni* or *Campylobacter coli*.

In other aspects, the present invention provides a method of producing a bioconjugate in a prokaryotic host cell, the method comprising the steps of:

- a. obtaining a prokaryotic host cell of the invention that produces a β -1,2 mannan polymer; and
- b. further introducing and expressing in the host cell:
 - i. a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are any amino acid except proline (optionally a polynucleotide sequence of the invention), and wherein the modified carrier protein further comprises an N-terminal bacterial signal sequence capable of transporting the modified carrier protein to the periplasmic side of the inner membrane of the prokaryotic host cell; and
 - ii. a nucleotide sequence encoding an oligosaccharyl transferase capable of producing a bioconjugate by transferring the β -1,2 mannan polymer from a lipid carrier to the modified carrier protein, wherein the oligosaccharyl transferase is PglB from *Campylobacter*, optionally from *Campylobacter jejuni* or *Campylobacter coli*.

In certain aspects, the modified carrier protein of the invention includes, but is not limited to, Sap2, Als3, detoxified Exotoxin A of *P. aeruginosa* (EPA), CRM197, Diphtheria toxoid, tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A of *S. aureus*, clumping factor B of *S. aureus*, *E. coli* FimH, *E. coli* FimHC, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* sat protein, the passenger domain of *E. coli* sat protein, *C. jejuni* AcrA, and a *C. jejuni* natural glycoprotein. In specific aspects, the modified carrier protein is a modified Sap2 protein of the invention.

In certain embodiments, the bacterial signal sequence is selected from, without limitation, FlgI, MalE, OmpA, and OmpC. In specific aspects, the bacterial signal sequence is FlgI. In certain aspects, the FlgI comprises an amino acid sequence of SEQ ID NO: 21. In particular embodiments, the bacterial signal sequence is removed from the modified carrier protein after the modified carrier protein is transported to the periplasmic side of the inner membrane of the prokaryotic host cell.

Analytical Methods

Various methods can be used to analyze the structural compositions and sugar chain lengths of the bioconjugates of the invention and to determine glycosylation site usage.

Hydrazinolysis can be used to analyze glycans. First, polysaccharides are released from their protein carriers by incubation with hydrazine according to the manufacturer's instructions (Ludger

Liberate Hydrazinolysis Glycan Release Kit, Oxfordshire, UK). The nucleophile hydrazine attacks the glycosidic bond between the polysaccharide and the carrier protein and allows release of the attached glycans. N-acetyl groups are lost during this treatment and have to be reconstituted by re-N-acetylation. The free glycans are purified on carbon columns and subsequently labeled at the reducing
5 end with the fluorophor 2-amino benzamide. See Bigge JC, Patel TP, Bruce JA, Goulding PN, Charles SM, Parekh RB, Anal Biochem 1995, 230(2):229-238. The labeled polysaccharides are separated on a GlycoSep-N column (GL Sciences) according to the HPLC protocol of Royle et al. See Royle L, Mattu TS, Hart E, Langridge JI, Merry AH, Murphy N, Harvey DJ, Dwek RA, Rudd PM, Anal Biochem 2002, 304(1):70-90. The resulting fluorescence chromatogram indicates the polysaccharide length and
10 number of repeating units. Structural information can be gathered by collecting individual peaks and subsequently performing MS/MS analysis. Thereby the monosaccharide composition and sequence of the repeating unit can be confirmed and additionally in homogeneity of the polysaccharide composition can be identified. Alternatively, high mass MS and size exclusion HPLC can be applied to measure the size of the complete bioconjugates.

15 Yield may be measured as carbohydrate amount derived from a liter of bacterial production culture grown in a bioreactor under controlled and optimized conditions. After purification of bioconjugate, the carbohydrate yields can be directly measured by either the anthrone assay or ELISA using carbohydrate specific antisera. Indirect measurements are possible by using the protein amount (measured by BCA, Lowry, or bardford assays) and the glycan length and structure to calculate a
20 theoretical carbohydrate amount per gram of protein. In addition, yield can also be measured by drying the glycoprotein preparation from a volatile buffer and using a balance to measure the weight.

Various methods can be used to analyze the conjugates of the invention including, for example, SDS-PAGE or capillary gel electrophoresis. Polymer length is defined by the number of repeat units that are linearly assembled. This means that the typical ladder like pattern is a
25 consequence of different repeat unit numbers that compose the glycan. Thus, two bands next to each other in SDS PAGE (or other techniques that separate by size) differ by only a single repeat unit. These discrete differences are exploited when analyzing glycoproteins for glycan size: the unglycosylated carrier protein and the bioconjugate with different polymer chain lengths separate according to their electrophoretic mobilities. The first detectable repeat unit number (n_1) and the
30 average repeat unit number ($n_{average}$) present on a bioconjugate are measured. These parameters can be used to demonstrate batch to batch consistency or polysaccharide stability, for example.

Glycosylation site usage may be quantified by, for example, glycopeptide LC-MS/MS: conjugates are digested with protease(s), and the peptides are separated by a suitable chromatographic method (C18, Hydrophilic interaction HPLC HILIC, GlycoSepN columns, SE HPLC, AE
35 HPLC), and the different peptides are identified using MS/MS. This method can be used with or without previous sugar chain shortening by chemical (smith degradation) or enzymatic methods.

Quantification of glycopeptide peaks using UV detection at 215 to 280nm allows relative determination of glycosylation site usage. In other embodiments, site usage may be quantified by size exclusion HPLC: Higher glycosylation site usage is reflected by an earlier elution time from a SE HPLC column. In yet other embodiments, site usage may be quantified by quantitative densitometry of purified
5 bioconjugates stained with Coomassie Brilliant Blue following sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Immunogenic Compositions and Vaccines

The conjugates (*e.g.* bioconjugate), of the invention are particularly suited for inclusion in
10 immunogenic compositions and vaccines.

Thus, in certain embodiments, the present invention provides an immunogenic composition comprising a conjugate or a bioconjugate of the invention. In certain aspects, the immunogenic composition additionally comprises a pharmaceutically acceptable excipient and/or carrier. In specific aspects, the present invention provides an immunogenic composition comprising a modified Sap2
15 protein of the invention, a conjugate of the invention, or a bioconjugate of the invention. In other aspects, the immunogenic composition additionally comprises a pharmaceutically acceptable excipient and/or carrier.

Immunogenic compositions comprise an immunologically effective amount of a modified Sap2 protein of the invention, or conjugate (*e.g.* bioconjugate) of the invention, as well as any other
20 components. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either as a single dose or as part of a series is effective for treatment or prevention. This amount varies depending on the health and physical condition of the individual to be treated, age, the degree of protection desired, the formulation of the vaccine and other relevant factors.

Pharmaceutically acceptable excipients and carriers are described, for example, in
25 Remington's Pharmaceutical Sciences, by E. W. Martin, Mack Publishing Co. Easton, PA, 5th Edition (1975). Pharmaceutically acceptable excipients can include a buffer, such as a phosphate buffer (*e.g.* sodium phosphate). Pharmaceutically acceptable excipients can include a salt, for example sodium chloride. Pharmaceutically acceptable excipients can include a solubilizing/stabilizing agent, for example, polysorbate (*e.g.* TWEEN 80). Pharmaceutically acceptable excipients can include a
30 preservative, for example 2-phenoxyethanol or thiomersal. Pharmaceutically acceptable excipients can include a carrier such as water or saline.

In additional embodiments, the present invention provides a method of making the immunogenic composition of the invention comprising the step of mixing a modified Sap2 protein of the invention, a conjugate of the invention, or a bioconjugate of the invention, with a pharmaceutically
35 acceptable excipient or carrier.

Preferably, an immunogenic composition of the invention is formulated as a vaccine for in vivo administration to a subject (*e.g.* human) wherein the individual components of the composition are formulated such that the immunogenicity of individual components is not impaired by other individual components of the composition (see above definition). Thus, in some embodiments, no (significantly) detrimental effect occurs to the saccharide antigen (in terms of protective efficacy) in the combination as compared to their administration in isolation. Preferably, an immunogenic composition of the invention is formulated as a vaccine for in vivo administration to a subject (*e.g.* human), which confers an antibody titre superior to the criterion for seroprotection for each antigenic component for an acceptable percentage of human subjects.

Thus, in additional embodiments, the present invention provides a vaccine comprising an immunogenic composition of the invention. In certain embodiments, the vaccine additionally comprises a pharmaceutically acceptable excipient or carrier. In other embodiments, the vaccine additionally comprises an adjuvant. In specific embodiments, the present invention provides a vaccine comprising an immunogenic composition of the invention and, optionally, a pharmaceutically acceptable excipient or carrier.

In specific embodiments, the present invention provides a *Candida albicans* vaccine comprising: (1) a modified Sap2 protein of the invention; (2) at least one *Candida albicans* saccharide antigen linked to said modified Sap2 protein; and, optionally, (3) a pharmaceutically acceptable carrier or adjuvant.

The term "adjuvant" refers to a compound that when administered in conjunction with or as part of an immunogenic composition of vaccine of the invention augments, enhances and/or boosts the immune response to modified Sap2 protein conjugate/bioconjugate, but when the compound is administered alone does not generate an immune response to the modified Sap2 protein conjugate/bioconjugate. Adjuvants can enhance an immune response by several mechanisms including, *e.g.* lymphocyte recruitment, stimulation of B and/or T cells, and stimulation of macrophages. Specific examples of adjuvants include, but are not limited to, aluminum salts (alum) (such as aluminum hydroxide, aluminum phosphate, and aluminum sulfate), 3 De-O-acylated monophosphoryl lipid A (MPL) (see United Kingdom Patent GB2220211), MF59 (Novartis), AS01 (GlaxoSmithKline), and saponins, such as QS21 (see Kensil *et al.* in Vaccine Design: The Subunit and Adjuvant Approach (eds. Powell & Newman, Plenum Press, NY, 1995); U.S. Pat. No. 5,057,540). In some embodiments, the adjuvant is Freund's adjuvant (complete or incomplete). Other adjuvants are oil in water emulsions (such as squalene or peanut oil), optionally in combination with immune stimulants, such as monophosphoryl lipid A (see Stoute *et al.* N. Engl. J. Med. 336, 86-91 (1997)). A suitable adjuvant is an adjuvant comprising an oil in water emulsion, wherein said oil in water emulsion comprises a metabolisable oil, a tocol and an emulsifier. Suitably, the metabolisable oil is squalene, the tocol is alpha-tocopherol and the emulsifying agent is polyoxyethylene sorbitan monooleate.

In some embodiments, the adjuvant includes an oil-in-water emulsion. For example, the oil-in-water emulsion can include an oil phase that incorporates a metabolisable oil, and an additional oil phase component, such as a tocol. The oil-in-water emulsion may also contain an aqueous component, such as a buffered saline solution (e.g., phosphate buffered saline). In addition, the oil-in-water emulsion typically contains an emulsifier. In some embodiments, the metabolizable oil is squalene. In certain embodiments, the tocol is alpha-tocopherol. In some embodiments, the emulsifier is a nonionic surfactant emulsifier (such as polyoxyethethylene sorbitan monooleate, TWEEN80™). In exemplary embodiments, the oil-in-water emulsion contains squalene and alpha tocopherol in a ratio which is equal or less than 1 (w/w).

10 In certain embodiments, the metabolisable oil in the oil-in-water emulsion may be present in an amount of 0.5-10mg. In some embodiments, the tocol in the oil-in-water emulsion may be present in an amount of 0.5 – 11 mg. In some embodiments, the emulsifying agent may be present in an amount of 0.4 – 4 mg.

15 In order for any oil in water composition to be suitable for human administration, the oil phase of the emulsion system has to comprise a metabolisable oil. The meaning of the term metabolisable oil is well known in the art. Metabolisable can be defined as 'being capable of being transformed by metabolism' (Dorland's Illustrated Medical Dictionary, W.B. Sanders Company, 25th edition (1974)). The oil may be any vegetable oil, fish oil, animal oil or synthetic oil, which is not toxic to the recipient and is capable of being transformed by metabolism. Nuts, seeds, and grains are common sources of vegetable oils. Synthetic oils are also part of this invention and can include commercially available oils such as NEOBEE® (caprylic/capric triglycerides made using glycerol from vegetable oil sources and medium-chain fatty acids (MCTs) from coconut or palm kernel oils) and others. A particularly suitable metabolisable oil is squalene. Squalene (2,6,10,15,19,23-Hexamethyl-2,6,10,14,18,22-tetracosahexaene) is an unsaturated oil which is found in large quantities in shark-liver oil, and in lower quantities in olive oil, wheat germ oil, rice bran oil, and yeast, and is a particularly preferred oil for use in this invention. Squalene is a metabolisable oil by virtue of the fact that it is an intermediate in the biosynthesis of cholesterol (Merck index, 10th Edition, entry no.8619).

25 Tocols are well known in the art and are described in EP0382271. Suitably the tocol is alpha-tocopherol or a derivative thereof such as alpha-tocopherol succinate (also known as vitamin E succinate). In certain embodiments, said tocol is suitably present in in an amount of 0.5-11 mg.

30 The oil in water emulsion further comprises an emulsifying agent. The emulsifying agent may suitably be polyoxyethylene sorbitan monooleate. In a particular embodiment the emulsifying agent may be Polysorbate® 80 (Polyoxyethylene (20) sorbitan monooleate) or Tween® 80. In certain embodiments, said emulsifying agent is suitably present in the adjuvant composition in an amount of 0.4-4mg.

Also provided is a method of making the immunogenic composition of the invention comprising the step of mixing the modified Sap2 protein or the conjugate (*e.g.* bioconjugate) of the invention with a pharmaceutically acceptable excipient and/or carrier and an adjuvant. Vaccine preparation is generally described in Vaccine Design ("The subunit and adjuvant approach" (eds Powell M.F. & Newman M.J.) (1995) Plenum Press New York).

The immunogenic compositions of the invention can be included in a container, pack, or dispenser together with instructions for administration. The immunogenic compositions or vaccines of the invention can be stored before use, *e.g.* the compositions can be stored frozen (*e.g.* at about -20°C or at about -70°C); stored in refrigerated conditions (*e.g.* at about 4°C); or stored at room temperature. The immunogenic compositions or vaccines of the invention may be stored in solution or lyophilized. In certain embodiments, the solution is lyophilized in the presence of a sugar such as sucrose, trehalose or lactose. In additional embodiments, the vaccines of the invention are lyophilized and extemporaneously reconstituted prior to use.

15 Administration and Dosage

Immunogenic compositions or vaccines of the invention may be used to protect or treat a subject (*e.g.* human), by means of administering said immunogenic composition or vaccine via systemic or mucosal route. These administrations may include injection via the intramuscular (IM), intraperitoneal, intradermal (ID) or subcutaneous (SC) routes; or via mucosal administration to the oral/alimentary, respiratory, genitourinary tracts.

In certain aspects, the immunogenic composition or vaccine of the invention is administered by the intramuscular delivery route. Intramuscular administration may be to the thigh or the upper arm. Injection is typically via a needle (*e.g.* a hypodermic needle), but needle-free injection may alternatively be used. A typical intramuscular dose is 0.5 ml.

In another aspects, the immunogenic composition or vaccine of the invention is administered by the intradermal administration. Human skin comprises an outer "horny" cuticle, called the stratum corneum, which overlays the epidermis. Underneath this epidermis is a layer called the dermis, which in turn overlays the subcutaneous tissue. The conventional technique of intradermal injection, the "mantoux procedure", comprises steps of cleaning the skin, and then stretching with one hand, and with the bevel of a narrow gauge needle (26 to 31 gauge) facing upwards the needle is inserted at an angle of between 10 to 15. Once the bevel of the needle is inserted, the barrel of the needle is lowered and further advanced whilst providing a slight pressure to elevate it under the skin. The liquid is then injected very slowly thereby forming a bleb or bump on the skin surface, followed by slow withdrawal of the needle.

In additional aspects, the immunogenic composition or vaccine of the invention is administered by the intranasal administration. Typically, the immunogenic composition or vaccine is administered

locally to the nasopharyngeal area, *e.g.* without being inhaled into the lungs. It is desirable to use an intranasal delivery device which delivers the immunogenic composition or vaccine formulation to the nasopharyngeal area, without or substantially without it entering the lungs. Suitable devices for intranasal administration of the vaccines according to the invention are spray devices. Suitable commercially available nasal spray devices include ACCUSPRAY™ (Becton Dickinson).

The amount of conjugate (*e.g.* bioconjugate) in each immunogenic composition or vaccine dose is selected as an amount which induces an immunoprotective response without significant, adverse side effects in typical vaccines. Such amount will vary depending upon which specific immunogen is employed and how it is presented. The content of a conjugate (*e.g.* a bioconjugate) will typically be in the range 1-100µg, suitably 5-50µg for the saccharide dose, *e.g.*, mannan dose.

Prophylactic and Therapeutic Uses

The present invention also provides an immunogenic composition of the invention, or the vaccine of the invention, for use in medicine. The use of an immunogenic composition of the invention in the manufacture of a medicament for the treatment or prevention of diseases caused by infection by *C. albicans* is also envisioned, as is a method of immunising a subject (*e.g.* a human) against disease caused by *C. albicans*, which method comprises administering to the subject an immunoprotective dose of an immunogenic composition of the invention.

In certain embodiments, the present invention provides a method of inducing immune response to a fungal infection in a subject (*e.g.* human), the method comprising administering to the subject a therapeutically or prophylactically effective amount of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention. In certain embodiments, the modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention can be used to induce an immune response against fungi. In certain aspects, the fungi is *Candida* species. In some aspects, the *Candida* species includes, but is not limited to, *Candida albicans*, *Candida auris*, *Candida guilliermondii*, *Candida lusitanaea* and *Candida tropicalis*, *Candida glabrata*, *Candida krusei*, and *Candida parapsilosis*. In specific embodiments, the *Candida* species is *Candida albicans*. In certain embodiments, said subject has fungal infection at the time of administration. In other embodiments, said subject does not have a fungal infection at the time of administration.

In specific embodiments, the present invention provides a method of inducing immune response to *Candida albicans* infection in a subject (*e.g.* human), the method comprising administering to the subject a therapeutically or prophylactically effective amount of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention.

The present invention also provides methods of treating and/or preventing a fungal infection in a subject comprising administering to the subject a conjugate (*e.g.* bioconjugate) of the invention. The conjugate (*e.g.* bioconjugate) may be in the form of an immunogenic composition or vaccine. Thus, in certain embodiments, the present invention provides a method for treatment or prevention of fungal infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention. The conjugate (*e.g.* bioconjugate) may be in the form of an immunogenic composition or vaccine. In other embodiments, the present invention provides a method for immunizing a subject against fungal infection, the method comprising administering to the subject an immunoprotective dose of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention. In still other embodiments, the present invention provides a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention for use in the manufacture of a medicament for the treatment or prevention of a disease caused by fungal infection in a subject (*e.g.* human). In certain aspects, the fungi is *Candida* species. In some aspects, the *Candida* species includes, but is not limited to, *Candida albicans*, *Candida auris*, *Candida guilliermondii*, *Candida lusitanaea* and *Candida tropicalis*, *Candida glabrata*, *Candida krusei*, and *Candida parapsilosis*. In specific embodiments, the *Candida* species is *Candida albicans*.

This, in specific embodiments, the present invention provides a method for treatment or prevention of *Candida albicans* infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention.

In other specific embodiments, the present invention provides a method for immunizing a subject against *Candida albicans* infection, the method comprising administering to the subject an immunoprotective dose of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention.

In specific embodiments, the present invention provides a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention for use in the manufacture of a medicament for the treatment or prevention of a disease caused by *Candida albicans* infection in a subject (*e.g.* human).

In other embodiments, the present invention provides a method of inhibiting adhesion of *Candida albicans* hyphae to vaginal epithelial cells in a subject (*e.g.* human), the method comprising

administering to the subject a therapeutically or prophylactically effective amount of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention.

5 In further embodiments, the present invention provides a method of mediating neutrophil killing of *Candida albicans* hyphae in a subject (*e.g.* human), the method comprising administering to the subject a therapeutically or prophylactically effective amount of a modified Sap2 protein of the invention, a conjugate of the invention, a bioconjugate of the invention, an immunogenic composition of the invention, or a vaccine of the invention.

Embodiments of the invention are further described in the subsequent numbered paragraphs:

10

1. A modified Secreted Aspartyl Proteinase 2 (Sap2) protein comprising amino acid residues 19-398 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1, modified in that the amino acid sequence comprises one or more consensus sequences comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.
- 15 2. The modified Sap2 protein of paragraph 1, wherein the protein further comprises a substitution at amino acid residue 274 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 20 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.
3. The modified Sap2 protein of paragraph 2, wherein the protein comprises an Aspartic Acid (D) to Asparagine (N) substitution at amino acid residue 274 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.
- 25 4. The modified Sap2 protein of any of paragraphs 2 and 3, wherein the substitution renders the Sap2 protein inactive.
5. The modified Sap2 protein of any of paragraphs 1 to 4, wherein the one or more consensus sequences have been added next to or substituted for one or more amino acids selected from 30 the group consisting of one or more amino acids between amino acid residues 19-24, one or more amino acids between amino acid residues 52-62, one or more amino acids between amino acid residues 93-107, one or more amino acids between amino acid residues 104-118,

- one or more amino acids between amino acid residues 142-144, one or more amino acids between amino acid residues 187-197, one or more amino acids between amino acid residues 215-225, one or more amino acids between amino acid residues 243-253, one or more amino acids between amino acid residues 255-265, one or more amino acids between amino acid residues 263-273, one or more amino acids between amino acid residues 320-330, one or more amino acids between amino acid residues 339-349, and one or more amino acids between amino acid residues 358-359 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.
- 5
- 10
- 15
- 20
- 25
- 30
6. The modified Sap2 protein of any of paragraphs 1 to 5, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 19 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.
 7. The modified Sap2 protein of any of paragraphs 1 to 6, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 57 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.
 8. The modified Sap2 protein of any of paragraphs 1 to 7, wherein the one or more consensus sequences have been added next to or substituted for one or more amino acids between amino acid residues 98-102 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.
 9. The modified Sap2 protein of any of paragraphs 1 to 8, wherein the one or more consensus sequences have been added next to or substituted for one or more amino acids between amino acid residues 109-113 of amino acid residues 19-398 of SEQ ID NO: 1 or at equivalent positions within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.
 10. The modified Sap2 protein of any of claims 1 to 9, wherein the one or more consensus sequences have been added next to or substituted for amino acid residue 220 of amino acid

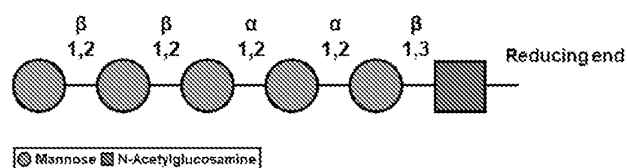
residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.

- 5 11. The modified Sap2 protein of any of paragraphs 1 to 10, wherein the modified Sap2 protein is from *Candida*.
12. The modified Sap2 protein of paragraph 11, wherein the *Candida* is selected from the group consisting of *Candida albicans*, *Candida auris*, *Candida guilliermondii*, *Candida lusitanaea*, *Candida tropicalis*, *Candida glabrata*, *Candida krusei*, and *Candida parapsilosis*.
- 10 13. The modified Sap2 protein of any of paragraphs 11 and 12, wherein the modified Sap2 protein is from *Candida albicans*.
14. The modified Sap2 protein of any of paragraphs 1 to 13, wherein the modified Sap2 protein further comprises at least one Fructose biphosphate aldolase-1 (Fba) peptide comprising an amino acid sequence of YGKDVKDLFDYAE (SEQ ID NO: 3) or an amino acid sequence at least 70%, 80%, 85%, 90%, or 92% identical to SEQ ID NO: 3.
- 15 15. The modified Sap2 protein of paragraph 14, wherein the Fba peptide is covalently linked to the modified Sap2 protein at amino acid residue 398 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, modified Sap2 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.
- 20 16. The modified Sap2 protein of any of paragraphs 1 to 15, wherein the one or more consensus sequences are selected from the group consisting of DQNAT (SEQ ID NO: 4) and DQNVT (SEQ ID NO: 5).
- 25 17. The modified Sap2 protein of any of paragraphs 14 to 16, wherein the modified Sap2 protein comprises at least one additional consensus sequence comprising an amino acid sequence of J-U-B-D/E-X-N-Z-S/T-J-U-B, wherein X and Z are independently any amino acid except proline and J, U and B independently comprise 1 to 5 naturally occurring amino acid residues, wherein the at least one consensus sequence has been added next to C-terminal amino acid residue of SEQ ID NO: 3 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 3.

18. The modified Sap2 protein of paragraph 17, wherein X is Q, Z is A, each of J and B comprises 1 to 5 Glycine (G) residues and U comprises 1 to 5 serine (S) residues.
19. The modified Sap2 protein of any of paragraphs 17 and 18, wherein the additional consensus sequence comprises an amino acid sequence of GSGGGDQ^NATGSGGG (SEQ ID NO: 7).
- 5 20. The modified Sap2 protein of any of paragraphs 17 to 19, wherein the additional consensus sequence comprises an amino acid sequence of GSGGGDQ^NATGSGGGHHHHHHHHHH (SEQ ID NO: 8).
21. The modified Sap2 protein of any of paragraphs 1 to 20 comprising an amino acid sequence of SEQ ID NO: 9.
- 10 22. A modified Sap2 protein comprising an amino acid sequence of SEQ ID NO: 10.
23. The modified Sap2 protein of any of paragraphs 1 to 22, wherein the modified Sap2 protein is glycosylated.
24. The modified Sap2 protein of any of paragraphs 1 to 23, wherein the modified Sap2 protein is N-glycosylated.
- 15 25. A conjugate comprising the modified Sap2 protein of any of paragraphs 1 to 22 and at least one saccharide antigen.
26. The conjugate of paragraph 25, wherein the modified Sap2 protein is covalently linked to the at least one saccharide antigen.
- 20 27. The conjugate of any of paragraphs 25 and 26, wherein the at least one saccharide antigen is selected from the group consisting of: *Citrobacter freundii* O antigen, *Agrobacterium* sp. Exopolysaccharide, O antigens of *E. coli*, *Salmonella* sp. O antigens, *Pseudomonas* sp., *Klebsiella* sp. O antigens, *Acinetobacter* O antigens, *Chlamydia trachomatis* O antigens, *Vibrio cholera* O antigens, *Listeria* sp. O antigens, *Legionella pneumophila* serotypes 1 to 15 O antigens, *Bordetella paraptussis* O antigens, *Burkholderia mallei* and *pseudomallei* O antigens, *Francisella tularensis* O antigens, *Campylobacter* sp. O antigens, capsular polysaccharides of *Clostridium difficile*, *Staphylococcus aureus* type 5 and 8, *Streptococcus pyrogenes*, *E. coli*, *Streptococcus agalacticae*, *Neisseria meningitidis*, *Candida* sp., *Candida albicans*, *Haemophilus influenza*, *Enterococcus faecalis* capsular polysaccharides type I-V, and
- 25

other surface polysaccharide structures, *e.g.* the *Borrelia burgdorferi* glycolipids, *Neisseria meningitidis* pilin O glycan and lipooligosaccharide (LOS), *Haemophilus influenza* LOS, *Leishmania major* lipophosphoglycan, tumor associated carbohydrate antigens, malaria glycosyl phosphatidylinositol, and *mycobacterium tuberculosis* arabinomannan.

- 5 28. The conjugate of any of paragraphs 25 and 26, wherein the at least one saccharide antigen is a fungal saccharide antigen.
29. The conjugate of any of paragraphs 25, 26 and 28, wherein the at least one saccharide antigen is a saccharide antigen of *Candida* species.
- 10 30. The conjugate of paragraph 29, wherein the *Candida* species is selected from the group consisting of *Candida albicans*, *Candida auris*, *Candida guilliermondi*, *Candida lusitanaea*, *Candida tropicalis*, *Candida glabrata*, *Candida krusei*, and *Candida parapsilosis*.
31. The conjugate of any of paragraphs 25 to 30, wherein the at least one saccharide antigen is a saccharide antigen of *Candida albicans*.
- 15 32. The conjugate of any of paragraphs 25 to 31, wherein the at least one saccharide antigen comprises a β -1,2 mannan polymer.
33. The conjugate of paragraph 32, wherein the β -1,2 mannan polymer comprises at least two β -1,2 linked mannose molecules.
34. The conjugate of paragraph 25 having the structure:



- 20 35. The conjugate of any of paragraphs 32 to 34, wherein the at least one saccharide antigen comprises the structure:
- β -D-Man β -(1 $\rightarrow$ 2)- β -D-Man β -(1 $\rightarrow$ 2)- α -D-Man β -(1 $\rightarrow$ 2)- α -D-Man β -(1 $\rightarrow$ 2)- β -D-Man β -(1 $\rightarrow$ 3)-x-D-GlcNAc.

36. The conjugate of any of paragraphs 25 to 31, wherein the at least one saccharide antigen comprises a β -1,2 mannan polymer.
37. The conjugate of paragraph 36, wherein the β -1,2 mannan polymer comprises at least two β -1,2 linked mannan molecules.
- 5 38. The conjugate of any of paragraphs 36 and 37, wherein the β -1,2 mannan polymer comprises two to fifty, two to forty, two to thirty, two to twenty, two to ten, two to nine, two to eight, two to seven, two to six, two to five, two to four, two to three, two, three to fifty, three to forty, three to thirty, three to twenty, three to ten, three to nine, three to eight, three to seven, three to six, three to five, three to four, three, four to fifty, four to forty, four to thirty, 10 four to twenty, four to ten, four to nine, four to eight, four to seven, four to six, four to five, four, five to fifty, five to forty, five to thirty, five to twenty, five to ten, five to nine, five to eight, five to seven, five to six, or five, six to forty, six to thirty, six to twenty, six to ten, six to nine, six to eight, six to seven, six, seven to fifty, seven to forty, seven to thirty, seven to twenty, seven to ten, seven to nine, seven to eight, or seven β -1,2 linked mannose molecules.
- 15 39. The conjugate of any of paragraphs 36 to 38, wherein the β -1,2 mannan polymer comprises at least 2, at least 3, at least 4, or at least 5 β -1,3 linked glucose molecules, optionally at least five consecutive β -1,2 linked mannan molecules, optionally five consecutive β -1,2 linked mannan molecules.
- 20 40. The conjugate of any of paragraphs 25 to 39, wherein the modified Sap2 protein is linked to the at least one saccharide antigen at one or more amino acid residues on the modified Sap2 protein, wherein the one or more residues are selected from the group consisting of one or more asparagine residues, one or more aspartic acid residues, one or more glutamic acid residues, one or more lysine residues, one or more cysteine residues, one or more tyrosine residues, one or more histidine residues, one or more arginine residues, one or more 25 tryptophan residues, one or more serine residues, and one or more threonine residues.
41. The conjugate of any of paragraphs 25 to 40, wherein the modified Sap2 protein is linked to the at least one saccharide antigen at one or more asparagine residues on the modified Sap2 protein.
42. The conjugate of any of paragraphs 25 to 41, wherein the conjugate is a bioconjugate.

43. A modified Sap2 protein of protein of *Candida albicans* comprising (or consisting of): (1) an amino acid sequence of SEQ ID NO: 9 or SEQ ID NO: 10; and (2) at least one saccharide antigen of *Candida*, wherein the at least one saccharide antigen is a β -1,2 mannan polymer consisting of at least five consecutive β -1,2 linked mannose molecules, and wherein the at least one saccharide antigen is linked to at least one of four asparagine residues at positions 45, 94, 215, and 415 of SEQ ID NO: 9 or at least one of four asparagine residues at positions 6, 55, 176, and 376 of SEQ ID NO: 10.
44. A polynucleotide sequence encoding the modified Sap2 protein of any of paragraphs 1 to 22.
45. A vector comprising the polynucleotide sequence of paragraph 44.
46. A host cell comprising: (1) one or more polynucleotide sequences that encode one or more heterologous glycosyltransferases; (2) a polynucleotide sequence that encodes a heterologous oligosaccharyl transferase; (3) a polynucleotide sequence that encodes a modified Sap2 protein according to any of paragraphs 1 to 22; and, optionally, (4) a polynucleotide sequence that encodes a polymerase.
47. The host cell of paragraph 46, wherein the host cell is *Escherichia coli*.
48. A method of producing a bioconjugate that comprises a modified Sap2 protein linked to at least one saccharide, the method comprising: (1) culturing the host cell of any of paragraphs 46 and 47 under conditions suitable for the production of proteins; and (2) isolating the bioconjugate produced by said host cell, optionally isolating the bioconjugate from a periplasmic extract from the host cell.
49. A bioconjugate produced by the method of paragraph 48, wherein said bioconjugate comprises at least one saccharide linked to a modified Sap2 protein of any of paragraphs 1 to 22.
50. An immunogenic composition comprising the modified Sap2 protein of any of paragraphs 1 to 24, the conjugate of any of paragraphs 25 to 41, or the bioconjugate of any of paragraphs 42 and 49.
51. A method of making the immunogenic composition of paragraph 50, the method comprising the step of mixing the modified Sap2 protein of any of paragraphs 1 to 24, the conjugate of

any of paragraphs 25 to 41, or the bioconjugate of any of paragraphs 42 and 49 with a pharmaceutically acceptable excipient or carrier.

52. A vaccine comprising the immunogenic composition of paragraph 50 and, optionally, a pharmaceutically acceptable excipient or carrier.

5 53. A *Candida albicans* vaccine comprising: (1) the modified Sap2 protein of any of paragraphs 1 to 22; (2) at least one *Candida albicans* saccharide linked to said modified Sap2 protein; and, optionally, (3) a pharmaceutically acceptable carrier or adjuvant.

10 54. A method for treatment or prevention of *Candida albicans* infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of the modified Sap2 protein of any of paragraphs 1 to 24 and 43, the conjugate of any of paragraphs 25 to 41, the bioconjugate of any of paragraphs 42 and 49, the immunogenic composition of paragraph 50, or the vaccine of any of paragraphs 52 and 53.

55. The method of paragraph 54, wherein the *Candida albicans* infection causes Recurrent Vulvovaginal Candidiasis (RVVC) in the subject.

15 56. A method for immunizing a subject against *Candida albicans* infection, the method comprising administering to the subject an immunoprotective dose of the modified Sap2 protein of any of paragraphs 1 to 24 and 43, the conjugate of any of paragraphs 25 to 41, the bioconjugate of any of paragraphs 42 and 49, the immunogenic composition of paragraph 50, or the vaccine of any of paragraphs 52 and 53.

20 57. A method for inducing immune response to *Candida albicans* infection in a subject, the method comprising administering to the subject a therapeutically or prophylactically effective amount of modified Sap2 protein of any of paragraphs 1 to 24 and 43, the conjugate of any of paragraphs 25 to 41, the bioconjugate of any of paragraphs 42 and 49, the immunogenic composition of paragraph 50, or the vaccine of any of paragraphs 52 and 53.

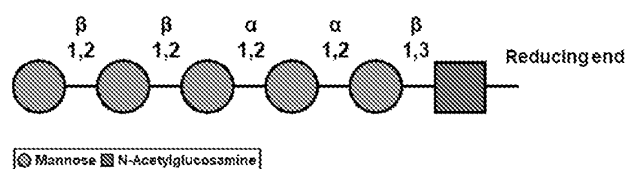
25 58. The method of any of paragraphs 54 to 57, wherein the subject is a human.

59. The modified Sap2 protein of any of paragraphs 1 to 24 and 43, the conjugate of any of paragraphs 25 to 41, the bioconjugate of any of paragraphs 42 and 49, the immunogenic

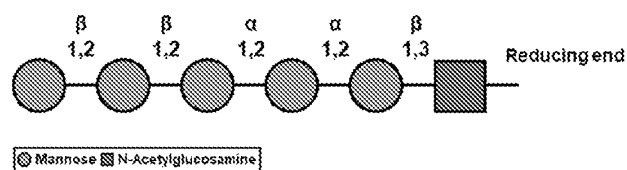
composition of paragraph 50, or the vaccine of any of paragraphs 52 and 53 for use in treatment or prevention of a disease caused by *Candida albicans* infection.

60. The modified Sap2 protein of any of paragraphs 1 to 24 and 43, the conjugate of any of paragraphs 25 to 41, the bioconjugate of any of paragraphs 42 and 49, the immunogenic composition of paragraph 50, or the vaccine of any of paragraphs 52 and 53 for use in manufacture of a medicament for treatment or prevention of a disease caused by *Candida albicans* infection.

61. A saccharide that is a β -1,2 mannan polymer comprising the structure:



62. The saccharide of paragraph 61 having the structure:



63. A saccharide comprising the structure: β -D-Man ρ -(1 $\rightarrow$ 2)- α -D-Man ρ -(1 $\rightarrow$ 2)- α -D-Man ρ -(1 $\rightarrow$ 2)- β -D-Man ρ -(1 $\rightarrow$ 3)-x-D-Glc ρ NAc.

64. The saccharide of paragraph 63 comprising the structure: β -D-Man ρ -(1 $\rightarrow$ 2)- β -D-Man ρ -(1 $\rightarrow$ 2)- α -D-Man ρ -(1 $\rightarrow$ 2)- α -D-Man ρ -(1 $\rightarrow$ 2)- β -D-Man ρ -(1 $\rightarrow$ 3)-x-D-Glc ρ NAc.

65. The saccharide of paragraph 64 which is linked to a lipid carrier.

66. The saccharide of paragraph 65, wherein the lipid carrier is undecaprenyl.

67. A conjugate comprising the saccharide of any one of paragraphs 61-64 linked to an asparagine residue of a modified carrier protein.

68. The conjugate of paragraph 67, wherein the asparagine residue is within a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.
69. The conjugate of any of paragraphs 67 and 68, wherein the modified carrier protein is selected from the group consisting of the modified Sap2 protein of any one of paragraphs 1 to 24, Sap2, Als3, detoxified Exotoxin A of *P. aeruginosa* (EPA), CRM197, Diphtheria toxoid, tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A of *S. aureus*, clumping factor B of *S. aureus*, *E. coli* FimH, *E. coli* FimHC, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli*/heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* sat protein, the passenger domain of *E. coli* sat protein, *C. jejuni* AcrA, and a *C. jejuni* natural glycoprotein.
70. The conjugate of any one of paragraphs 67 to 69, wherein the carrier protein is the modified Sap2 protein of any one of paragraphs 1 to 24.
71. The conjugate of any one of paragraphs 67 to 70, wherein the conjugate is a bioconjugate.
72. A host cell comprising:
- i. a nucleotide sequence encoding one or more first heterologous glycosyltransferase(s) capable of synthesizing a β -1,2 mannan polymer;
 - ii. a nucleotide sequence encoding a second heterologous glycosyltransferase which is eukaryotic and is capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain;
 - iii. a nucleotide sequence encoding a heterologous oligosaccharyl transferase; and
 - iv. optionally, a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.
73. The host cell of paragraph 72, wherein the one or more heterologous glycosyltransferases of i. comprises WbaD, WbaC, and WbaB from *Citrobacter*, optionally from *Citrobacter freundii*, optionally *Citrobacter freundii* P079F I.
74. The host cell of paragraph 73, wherein the one or more heterologous glycosyltransferase of i. has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of WbaD of *Citrobacter freundii* P079F I comprising SEQ ID NO: 11.

75. The host cell of any of paragraphs 73 and 74, wherein the one of more heterologous glycosyltransferase(s) of i. has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of WbaC of *Citrobacter freundii* P079F I comprising SEQ ID NO: 12.
- 5 76. The host cell of any one of paragraphs 73 to 75, wherein the one of more heterologous glycosyltransferase(s) of i. has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of WbaB of *Citrobacter freundii* P079F I comprising SEQ ID NO: 13.
- 10 77. The host cell of any one of paragraphs 73 to 76 further comprising a heterologous translocase, wherein the heterologous translocase has an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of Wzx of *Citrobacter freundii* P079F I comprising SEQ ID NO: 14.
- 15 78. The host cell of any one of paragraphs 73 to 77, wherein the eukaryotic glycosyltransferase of capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain ii. is Bmt3, optionally of *C. albicans*.
79. The host cell of paragraph 78, wherein the Bmt3 comprises an amino acid sequence at least 80%, 90%, 95%, 98%, or 99% identical to SEQ ID NO: 15.
80. The host cell of any one of paragraphs 73 to 79, wherein the oligosaccharyltransferase of iii. is a PglB
- 20 81. The host cell of paragraph 80, wherein the PglB is optionally from *Campylobacter*, optionally from *Campylobacter jejuni* or *Campylobacter coli*, optionally an evolved PglB comprising an amino acid sequence of SEQ ID NO: 16.
- 25 82. The host cell of any one of paragraphs 73 to 81, wherein the modified carrier protein of iv. is selected from the group consisting of Sap2, Als3, detoxified Exotoxin A of *P. aeruginosa* (EPA), CRM197, Diphtheria toxoid, tetanus toxoid, detoxified hemolysin A of *S. aureus*, clumping factor A of *S. aureus*, clumping factor B of *S. aureus*, *E. coli* FimH, *E. coli* FimHC, *E. coli* heat labile enterotoxin, detoxified variants of *E. coli* heat labile enterotoxin, Cholera toxin B subunit (CTB), cholera toxin, detoxified variants of cholera toxin, *E. coli* sat protein, the passenger domain of *E. coli* sat protein, *C. jejuni* AcrA, and a *C. jejuni* natural glycoprotein.

83. The host cell of paragraph 82, wherein the modified carrier protein is the modified Sap2 protein of any one of paragraphs 1 to 24.
84. The host cell of any one of paragraphs 73 to 83, wherein said host cell is an *Escherichia* species, *Shigella* species, *Klebsiella* species, *Xhantomonas* species, *Salmonella* species, *Yersinia* species, *Lactococcus* species, *Lactobacillus* species, *Pseudomonas* species, *Corynebacterium* species, *Streptomyces* species, *Streptococcus* species, *Staphylococcus* species, *Bacillus* species, or a *Clostridium* species.
85. The host cell of paragraph 84, wherein said host cell is an *E. coli* species.
86. A method of producing a glycoconjugate comprising a modified carrier protein and a β -1,2 mannan, wherein said method comprises the steps of i) culturing the host cell of any one of paragraphs 73 to 85 under conditions suitable for the production of proteins, ii) harvesting the culture to produce a harvested culture, and iii) isolating the glycoconjugate from the culture.
87. The method of paragraph 86, wherein the conditions suitable for the production of the glycoconjugate includes the addition of GDP-mannose to the culture medium.
88. The method of paragraph 86, wherein GDP-mannose is added to the harvested culture.
89. A method of producing a bioconjugate in a prokaryotic host cell, the method comprising the steps of:
- obtaining a prokaryotic host cell of any of paragraphs 73 to 85 that produces a β -1,2 mannan polymer; and
 - further introducing and expressing in the host cell:
 - a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are any amino acid except proline, and wherein the modified carrier protein further comprises an N-terminal bacterial signal sequence capable of transporting the modified carrier protein to the periplasmic side of the inner membrane of the prokaryotic host cell; and
 - a nucleotide sequence encoding an oligosaccharyl transferase capable of producing a bioconjugate by transferring the β -1,2 mannan polymer from a lipid carrier to the modified

carrier protein, wherein the oligosaccharyl transferase is PglB from *Campylobacter*, optionally from *Campylobacter jejuni* or *Campylobacter coli*.

90. The method of paragraph 89, wherein the bacterial signal sequence is selected from the group consisting of: FlgI, MalE, OmpA, and OmpC.
- 5 91. The method of paragraph 90, wherein the bacterial signal sequence is FlgI.
92. The method of paragraph 91, wherein the FlgI comprises an amino acid sequence of SEQ ID NO: 21.
93. The method of any of paragraphs 90 to 92, wherein the bacterial signal sequence is removed from the modified carrier protein after the modified carrier protein is transported to the periplasmic side of the inner membrane of the prokaryotic host cell.
- 10 94. The method of any of paragraphs 90 to 93, wherein the modified carrier protein is the modified Sap2 protein of any of paragraphs 1 to 24.
95. The modified Sap2 protein of any of paragraphs 14 to 16, wherein the modified Sap2 protein comprises at least one additional consensus sequence comprising an amino acid sequence of J-U-B-D/E-X-N-Z-S/T-J-U-B, wherein X and Z are independently any amino acid except proline and J, U and B independently comprise 1 to 5 naturally occurring amino acid residues, wherein the at least one additional consensus sequence has been added next to C-terminal amino acid residue of SEQ ID NO: 3 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 3.
- 15 96. A method of inhibiting adhesion of *Candida albicans* hyphae to vaginal epithelial cells in a subject (*e.g.* human), the method comprising administering to the subject a therapeutically or prophylactically effective amount of the modified Sap2 protein of any of paragraphs 1 to 24 and 43, the conjugate of any of paragraphs 25 to 41, the bioconjugate of any of paragraphs 42 and 49, the immunogenic composition of paragraph 50, or the vaccine of any of paragraphs 52 and 53.
- 20 97. A method of mediating neutrophil killing of *Candida albicans* hyphae in a subject (*e.g.* human), the method comprising administering to the subject a therapeutically or prophylactically effective amount of the modified Sap2 protein of any of paragraphs 1 to 24
- 25

and 43, the conjugate of any of paragraphs 25 to 41, the bioconjugate of any of paragraphs 42 and 49, the immunogenic composition of paragraph 50, or the vaccine of any of paragraphs 52 and 53.

5 98. A host cell comprising: (1) one or more polynucleotide sequences that encode one or more heterologous glycosyltransferases; (2) a polynucleotide sequence that encodes a heterologous oligosaccharyl transferase; (3) a polynucleotide sequence that encodes a modified Sap2 protein according to any of paragraphs 1 to 22 (optionally a polynucleotide sequence of paragraph 44); and, optionally, (4) a polynucleotide sequence that encodes a polymerase.

99. A host cell comprising:

10 i. a nucleotide sequence encoding one or more first heterologous glycosyltransferase(s) capable of synthesizing a β -1,2 mannan polymer;

ii. a nucleotide sequence encoding a second heterologous glycosyltransferase which is eukaryotic and is capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain;

15 iii. a nucleotide sequence encoding a heterologous oligosaccharyl transferase; and optionally, a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline, optionally a polynucleotide sequence of paragraph 44.

20 100. A method of producing a bioconjugate in a prokaryotic host cell, the method comprising the steps of:

a. obtaining a prokaryotic host cell of of any of paragraphs 72 to 84, 98 and 99 that produces a β -1,2 mannan polymer; and

b. further introducing and expressing in the host cell:

25 i. a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are any amino acid except proline (optionally a polynucleotide sequence of paragraph 44), and wherein the modified carrier protein further comprises an N-terminal bacterial signal sequence capable of transporting the modified

30 carrier protein to the periplasmic side of the inner membrane of the prokaryotic host cell; and

- ii. a nucleotide sequence encoding an oligosaccharyl transferase capable of producing a bioconjugate by transferring the β -1,2 mannan polymer from a lipid carrier to the modified carrier protein, wherein the oligosaccharyl transferase is PglB from *Campylobacter*, optionally from *Campylobacter jejuni* or *Campylobacter coli*.

All publications mentioned herein are herein incorporated by reference in their entirety. It is to be understood that the term "or," as used herein, denotes alternatives that may, where appropriate, be combined; that is, the term "or" includes each listed alternative separately as well as their combination. As used here, unless the context clearly dictates otherwise, references to the singular, such as singular forms "a," "an," and "the," include the plural, and references to the plural include the singular.

In order that this invention may be better understood, the following examples are set forth. These examples are for purposes of illustration only, and are not to be construed as limiting the scope of the invention in any manner.

15

EXAMPLES

Material and methods

Engineering of proSap2 for glycosylation with antigenic glycans

In order to predict suitable positions for insertion of glycosites, the crystal structure of proSap2 protein of the invention having an amino acid sequence of SEQ ID NO: 17 was analyzed using various software.

SEQ ID NO: 17 – Wild type intermediate Sap2 ("proSap2") sequence from *Candida albicans*:

25 TPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVT LHNEQV TYAADITVGSNNQKLN VIVDTGSS
DLWVPDVNVDCQVTYS DQTADFCKQKGTYPDPSGSSASQDLNTPFKIGYGDGSSSQGTLYKDTVGF GGVS IKNQ
VLADV DSTSIDQILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSLYLN SPDAA TGQII FGGVDNAKYSGSLI
ALPVTSDRELRLISLGSVEVSGKTINTDNVDVLVDSGTTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLS
30 GDVVFNF SKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTS
ASSISALT

For engineering proSap2 for glycosylation, the fragment corresponding to the proenzyme (residues 19-398) and comprising mutation D218N (second catalytic aspartate, numbering as in the mature Sap2 protein (mSap2; SEQ ID NO: 88)) was used. In total, 23 positions were selected for insertion of the consensus sequence for glycosylation i.e. glycosite (D/E-X-N-Z-S/T) by site directed mutagenesis. During the design of glycosite variants, for some of the positions slight variation of the substituted region or introduced consensus sequence was applied, yielding 37 mutants in total.

35

ProSap2 variants comprising single glycosites were tested for glycosylation with *Klebsiella pneumoniae* O5 antigen. The best performing glycosites were combined to generate variants comprising between two to six glycosites in total. The six positions that were selected for combinations were: 19 (N-term proSap2), 57 (N-term mSap2 = behind the propeptide), 98-102 (Mut3a), 109-113 (Mut4d), 220 (Mut8a), 398 (C-terminus), where the numbering corresponds to the residues of the native *Candida albicans* Sap2 sequence (SEQ ID NO: 1).

Glycosylation tests with engineered proSap2 comprising one or more glycosites

Modified proSap2 proteins comprising a single inserted glycosite were tested for *in vivo* glycosylation efficiency using *Klebsiella pneumoniae* O5 antigen. For the data set presented in this work, an *E. coli* W3110-derivative strain was used, which included the deletion of the LPS-O antigen ligase *waaL* and contained the cluster of genes for the biosynthesis of *Klebsiella pneumoniae* O5 glycan replacing the native O antigen cluster *rfaO16*. The *E. coli* strain producing KpO5 glycan was transformed with a pEC415 plasmid carrying a modified proSap2 protein and a plasmid expressing PglB. To prepare a pre-culture, 5 ml TB medium containing 10 mM MgCl₂ and appropriate antibiotics was inoculated with a streak of colonies from the transformation plate and grown at 37°C overnight. The pre-culture was used to inoculate 50 ml of supplemented TB medium in a shake flask to give a starting OD₆₀₀ = 0.1. The cultures were grown at 37°C, with 200 rpm shaking until reaching OD₆₀₀ = 0.8-1 and then induced by addition of 0.001% arabinose (proSap2) and 0.1 mM IPTG (PglB). The expression and glycosylation of modified proSap2 proteins was continued at 25°C overnight.

The selection criteria for modified proSap2 proteins with single glycosite included the total expression level and the level of produced glycoconjugate, the later indicating suitability of glycosite position for modification by PglB.

Periplasmic extract preparation

The amount of cells from overnight cultures corresponding to OD₆₀₀ = 60 (measured using a spectrophotometer) was harvested by centrifugation. The cell pellets were resuspended in 1.5 ml of lysis buffer (30 mM Tris-HCl pH 8.5, 1 mM EDTA (ethylenediaminetetraacetic acid), 20% sucrose) and lysozyme was added to a final concentration of 1 mg/ml. The suspensions were incubated with slight shaking for 25 minutes at 4°C and then centrifuged at 16,000 rcf for 10 min. After centrifugation, the supernatant corresponding to periplasmic extract (PPE) was transferred to a fresh tube.

Enrichment of periplasmic extract by immobilized metal affinity chromatography (IMAC)

In order to enrich periplasmic extracts with modified proSap2 proteins and allow more direct read-out by SDS-PAGE, His-tagged modified proSap2 proteins were purified using one-step purification on Ni-NTA (Nickel Nitrilo-triacetic Acid) agarose. 1ml of PPE was mixed with 200 µl of pre-equilibrated

Ni-NTA slurry and incubated with slight shaking for 30 min. After that, the resin was washed and the bound protein eluted with elution buffer (30 mM Tris pH 8.0, 500 mM imidazole, 50 mM NaCl). The IMAC enriched PPE was analysed by SDS-PAGE. "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4", *Nature*, 227 (5259): 680–685. Bibcode:1970 Natur.227..680L. doi:10.1038/227680a0. ISSN 0028-0836. PMID 5432063). Unglycosylated proSap2 proteins and glycoconjugates (*i.e.* modified proSap2 proteins linked to one or more polysaccharide chains) glycosylated at one or more positions were detected on the gel by Coomassie staining (Fazekas de St. Groth, S.; Webster, R. G.; Datyner, A. (1963). "Two new staining procedures for quantitative estimation of proteins on electrophoretic strips". *Biochimica et Biophysica Acta*. 71: 377–391. doi:10.1016/0006-3002(63)91092-8. PMID 18421828).

The results of these experiments are described in FIGS. 3-13 and the following Examples.

Example 1: SDS PAGE analysis of Modified proSap2 protein-single glycosite variants purified from PPE by IMAC

SDS-PAGE analysis (FIG. 3) was carried out on IMAC enriched periplasmic extract of *E.coli* strains producing KpO5 polysaccharide and expressing PglB and modified proSap2 proteins with a glycosite D/E-X-N-Z-S/T introduced at specific positions into SEQ ID NO: 17, as shown below in Tables 1 and 2:

Table 1:

Modified Protein Name	SEQ ID NO:	Residues in proSap2 replaced/modified by glycosite (relative to wild type <i>Candida albicans</i> proSap2 sequence of SEQ ID NO: 1)	Lane (Fig. 3A)
N	44	19	1
N3C	45	57	2
C	46	398	3
Mut1a	47	66 - 69	4
Mut1b	48	66 - 69	5
Mut2a	49	77 - 80	6
Mut2b	50	77 - 80	7
Mut2c	51	80	8
Mut3a	52	98 - 102	9
Mut3b	53	98 - 103	10
Mut4a	54	104 - 108	11
Mut4c	55	107 - 111	12

Modified Protein Name	SEQ ID NO:	Residues in proSap2 replaced/modified by glycosite (relative to wild type <i>Candida albicans</i> proSap2 sequence of SEQ ID NO: 1)	Lane (Fig. 3A)
Mut4d	56	109 - 113	13
Mut4e	57	110 - 111	14
Mut5a	58	127 - 129	15
Mut5b	59	127	16
Mut6a	60	142 - 144	17
Mut6b	61	143 - 144	18
Mut7	62	192	19

Table 2:

Modified Protein Name	SEQ ID NO:	Residues in proSap2 replaced/modified by glycosite (relative to wild type <i>Candida albicans</i> proSap2 sequence of SEQ ID NO: 1)	Lane (Fig. 3B)
Mut8a	63	220	20
Mut8b	64	219	21
Mut9	65	237 - 239	22
Mut10	66	248	23
Mut11	67	255 - 256	24
Mut12	68	261	25
Mut13	69	268	26
Mut14a	70	295 - 299	27
Mut14b	71	298 - 302	28
Mut14c	72	300 - 304	29
Mut14d	73	303 - 306	30
Mut14e	74	302 - 303	31
Mut15	75	315	32
Mut16	76	325	33
Mut17	77	344	34
Mut18	78	358 - 359	35
Mut19	79	377 - 379	36
Mut20	80	391 - 393	37

The protein bands in FIGS. 3A and 3B correspond to the unglycosylated modified proSap2 protein ("uCarrier"), and to KpO5-modified proSap2 bioconjugates ("Conjugate") with one occupied glycosite.

As shown in FIGS. 3A and 3B, modified proSap2 proteins with single glycosite showed variability in their expression. In certain constructs (e.g. Mut14) the expression was almost completely lost, while for other ones the expression was similar to the ex[ression of the wild type (wt) proSap2 protein (SEQ ID NO: 17). Surprisingly, the proSap2 "C" variant (SEQ ID NO: 46) showed expression level superior to the wt.

The proSap2 "C" variant (SEQ ID NO: 46) was also the one with the highest degree of glycosylation (FIG. 3A, lane 3). Several other variants showed a high level of glycosylation, such as N (SEQ ID NO: 44), Mut3a (SEQ ID NO: 52), Mut4d (SEQ ID NO: 56), and Mut8a (SEQ ID NO: 63) (FIG. 3A, lanes 9, 13, FIG. 3B lane 20, respectively) and they were selected for glycosite combinations. The modified proSap2 proteins used in this analysis comprised a histine tag, however a skilled person will recognize that modified proSap2 proteins with the histidine tag removed could also be used for the same analysis. Thus, in certain embodiments, the invention provides for modified Sap2 or modified proSap2 proteins with the histidine tag removed.

Example 2: Analysis of the glycosylation of proSap2 carrying combined glycosites and the Fba peptide

The best performing glycosites were combined to generate variants comprising between two to six glycosites in total. The six positions that were selected for combinations were: 19 (N-term proSap2), 57 (N-term mSap2 = behind the propeptide), 98-102 (Mut3a), 109-113 (Mut4d), 220 (Mut8a), 398 (C-terminus), where the numbering corresponds to the residues of the native *Candida albicans* Sap2 sequence (SEQ ID NO: 1). Additionally, in certain embodiments, the Fba peptide was introduced to selected variants at the C-terminal of the protein, followed by the "C" glycosite. SDS-PAGE analysis (FIG. 4) was carried out on IMAC enriched periplasmic extract of *E.coli* strains producing KpO5 polysaccharide and expressing PglB and proSap2 variant proteins carrying the following combination of glycosites (Table 3):

Table 3:

Modified Protein Name	No. of engineered glycosites	Position of engineered glycosites	Residues in proSap2 replaced/modified by glycosite (relative to wild type <i>Candida</i>	Includes Fba peptide at the C-terminal (upstream	SEQ ID NO:	Lane (FIG. 4)
-----------------------	------------------------------	-----------------------------------	--	--	------------	---------------

			<i>albicans</i> proSap2 sequence of SEQ ID NO: 1)	of the C glycosite)?		
MutC	1	C	398	No	46	1
MutN3C-C	2	N3C and C	57 and 398	No	81	2
MutN3C-8a-C	3	N3C, 8a, and C	57, 220, and 398	No	82	3
MutN3C-3a- 8a-C	4	N3C, 3a, 8a, and C	57, 98-102, 220, and 398	No	83	4
Mut8a-fba-C	2	8a and C	220 and 398	Yes	84	5
MutN3C-fba-C	2	N3C and C	57 and 398	Yes	85	6
MutN3C-8a- fba-C	3	N3C, 8a, and C	57, 220 and 398	Yes	86	7
MutN3C-3a- 8a-fba-C	4	N3C, 3a, 8a, and C	57, 98-102, 220, and 398	Yes	87	8

As shown in FIG. 4, modified proSap2 proteins with combined glycosites were all expressed without any significant loss in yield. All the introduced glycosites were at least partially glycosylated, with species with up to four glycosites occupied, resulting in only minimal unglycosylated protein left and maximizing the sugar / protein ratio. The addition of the Fba peptide did not influence the expression level or the glycosylation efficiency. The modified proSap2 variant, MutN3C-3a-8a-fba-C (SEQ ID NO: 87) was selected as the optimal protein carrier. The modified proSap2 proteins used in this analysis comprised a histidine tag, however a skilled person will recognize that modified proSap2 proteins with the histidine tag removed could also be used for the same analysis. Thus, in certain embodiments, the invention provides for modified Sap2 or modified proSap2 proteins with the histidine tag removed.

Example 3: Production of the proSap2-Fba- β -1,2-mannan conjugate

Construction of the Modified proSap2-Fba- β -1,2-mannan-producing strain

An *E. coli* K12 W3110-derivative strain was constructed by subsequent replacements of the targeted gene clusters with the gene of interest, if any, followed by an FRT sites-flanked selection marker via λ -Red homologous recombination followed by FLP recombinase-catalysed marker removal as described (TE Kuhlman and EC Cox. Nucleic Acids Res. 2010 Apr; 38(6): e92). Seven homologous recombination / marker removal steps were carried out, removing genomic sequences of i. LPS-O antigen ligase *waaL* (GenBank NC_007779 position 3'842'208 to 3'843'467), ii. O16 O-antigen cluster

(*rfb* or *wb*, GenBank NC_007779 position 2'114'113 to 2'103'814), iii. colanic acid cluster (*wca*, GenBank NC_007779 position 2'138'241 to 2'118'033), iv. ECA cluster retaining *wecA* (*wec*, GenBank NC_007779 position 3'666'604 to 3'656'725), v. O16 *wzz2* or *cld* (GenBank NC_007779 position 2'099'458 to 2'100'438), vi. *gtrABS* or *yfdGHI* (GenBank NC_007779 position 2'473'301 to 2'475'908),
 5 vii. *araBA* (GenBank NC_007779 position 66'835 to 70'048). Moreover, two copies of the genes *wzx*, *wbaB*, *wbaC*, *wbaD* from *Citrobacter freundii* P079F I (GenBank QFTZ01000001.1 387'158 to 393'257) were introduced in the O16 O-antigen cluster locus (*rfb* or *wb*, GenBank NC_007779 position 2'114'113 to 2'103'814) and replacing the gene *yeaS* (GenBank NC_007779 position 1'881'835 to 1'882'473).

A first expression plasmid comprising the gene for the mannosyltransferase Bmt from *Candida*
 10 *albicans* (GenBank XP_717972.1), fused with a periplasmic signal peptide at the N-terminal and codon optimized for *E. coli*, followed by the *E. coli* genes *manC* and *manB* (GenBank: NC_007779 position 2'123'746 to 2'126'657) for the enhanced biosynthesis of GDP-mannose, in a pEC415 backbone (Schulz et al. *J Biol Chem.* 1998 Aug; 281(5380):1197-200) was constructed. A second expression plasmid encoding the selected proSap2-Fba variant in a plasmid backbone with trimethoprim
 15 resistance and BBR ori (medium copy), under arabinose promoter control, was constructed. A third expression plasmid encoding a sequence of *pglB* obtained by directed evolution in a pEXT21 backbone (Dykxhoorn DM et al. *Gene* 1996;177(1-2):133-6.), was constructed. All the plasmids were created via one or more steps of classical restriction cloning and Gibson assembly (Gibson, D.G., et al. (2009) *Nat. Methods* 6, 343-345) starting from synthetic DNA templates.

20 The conjugate-producing strain was obtained by transforming the engineered strain with the three plasmids. The role of each gene in the production of the bioconjugate, as well as the structure of the glycan are schematized in FIG. 2.

Production and analysis of the Modified proSap2-β-1,2-mannan bioconjugate

The conjugate-producing strain was grown in a fed-batch 10-L bioreactor in a buffered rich
 25 medium at 37°C and pH 7. When the OD_{600nm} of the culture showed a value of 20-25, the temperature was changed to 25°C, the bioconjugate production was induced with IPTG and arabinose, and a rich medium feed was started and stopped 24 hours after the induction. GDP-Mannose was then added to the fermenter vessel for a span of 40 minutes to a final concentration of 50 μM. Cells were harvested by centrifugation 26 hours after induction and washed and resuspended in TBSE buffer. An
 30 osmotic shock protocol was applied in which the cells were diluted in a 5-fold volume of H₂O for 1 hour, releasing the content of the periplasm in the supernatant. Cells were separated from the supernatant by centrifugation, and cell debris were removed by filtration through 0.45 and 0.2 μm filters. Tangential Flow Filtration (TFF) with a 10 kDa cutoff was applied to exchange to an acidic buffer (10mM Citric Acid 40mM NaCl pH4.0). A cation exchange chromatography was applied, leading

to an almost complete separation of the bioconjugate from the impurity. The desired fractions were pooled, diluted to a pH7 buffer, and the bioconjugate was further polished through a hydrophobic interaction column. The bioconjugate-containing fractions were pooled and TFF was employed to exchange the buffer to 10 mM NaPO₄, 150 mM NaCl, pH 6.5 and to concentrate to the desired volume.

- 5 As shown in FIG. 5, the purified bioconjugate was analysed by SDS-PAGE gels followed by i. coomassie blue staining (FIG. 5A), ii. anti-proSap2 immunoblot (against N-terminal "pro-peptide" region; FIG. 5B), iii, anti-Sap2 immunoblot (against peptide in the middle of the protein sequence; FIG. 5C), iv. anti-Fba immunoblot (FIG. 5D), and iv. anti-mannan immunoblot (FIG. 5E).

10 Example 4: Structural determination of proSap2 produced in *E. coli* via crystallography

Detoxified proSap2 was periplasmically overexpressed in *E. coli*, purified and concentrated using Amicon concentrator (6 ml, 10 kDa CO) at 3000 rpm to a concentration of 10 mg/mL. Crystallization plates were setup on using a TTP Labtech Mosquito LCP robot; the drop size was 200 nl protein + 200 nl precipitant solutions; employed crystallization screens were SG1, PACT, Morpheus and PGA (Molecular Dimensions), and tested temperatures were 20°C and 4°C. Several conditions allowed formation of crystals and tested for diffraction. Data was collected at the X06DA beamline at the Swiss Light Source.

- 15 The crystals diffracted to high resolution. The structure of proSap2 was determined by the molecular replacement method using the structure of mSap2 (mature Sap2, lacking the propeptide) from *C. albicans* as a search model (PDB ID:1EAG).

The structure was refined to 1.7 Å resolution. High quality electron density map allowed building of 362 residues of proSap2 out of 380; 18 residues of the propeptide could not be built suggesting flexible structure of those regions. The remaining part of the propeptide was modelled with an available AlphaFold model of proSap2. The obtained structure was superimposed to the structure of mSap2 form *C. albicans* as shown in Fig. 6.

25 Conclusion:

E. coli-produced proSap2 folds as the protein from Candida, resulting in an extremely similar structure. The propeptide present in proSap2 used for bioconjugation only marginally covers the active site, allowing potential antibody to access it.

30

Example 5: Circular dichroism (CD) spectroscopy characterization of wild-type proSap2 protein and modified proSap2-β-1,2-mannan bioconjugate

- The same batch of detoxified wild-type proSap2 analysed by crystallography and a modified Sap2 protein of the invention (modified proSap2-Fba-β-1,2-mannan bioconjugate, where proSap2 is the MutN3C-8a-fba-C variant) were analysed by near and far UV CD spectroscopy in order to obtain
- 35

information on the secondary and tertiary structure characteristics of the proteins and to compare them.

For near-UV CD, a Chirascan Q100 CD spectrometer (Applied Photophysics Ltd., UK) was used. The near-UV CD spectra in the wavelength range of 240-350 nm were recorded as co-addition of ten consecutive scans at a scanning speed of 30 nm min⁻¹ by applying the following settings for bandwidth: 1 nm and wavelength step size: 0.5 nm, and by using a 1-cm path length flow-through cuvette (Applied Photophysics Ltd., UK). For far-UV CD, a Chirascan Q100 CD spectrometer (Applied Photophysics Ltd., UK) was used for far-UV CD spectroscopy measurements. The far-UV CD spectra in the wavelength range of 190-260 nm were recorded as co-addition of ten consecutive scans at a scanning speed of 60 nm min⁻¹ by applying the following settings for bandwidth: 1 nm and wavelength step size: 1 nm, and by using a 0.01-cm path length flow-through cuvette (Applied Photophysics Ltd., UK). The formulation buffer was used as a blank. The mean residue ellipticities were extrapolated mathematically and plotted. For secondary structure element content estimation, the far-UV CD spectra were analyzed by using CDNN software (version 2.1) with 33 untrained neural networks and a database containing secondary structure data for the wavelength range of 178-260 nm from 29 soluble proteins (SP29 database, compiled by Johnson et al., 1981 and 1988) and four additional components, i.e., carbopeptidase A, polyglutamic acid, rubredoxin and trypsin. The retrieved data are reported in FigS. 7A (near-UV CD), 7B (far-UV CD) and 7C (secondary structure element content after analysis with "CDNN" software (Applied Photophysics Ltd) for deconvolution of CD spectroscopy).

20 Conclusion:

The analysis show that both proSap2 (as used for crystallography) and the proSap2-Fba- β -1,2-mannan bioconjugate share extremely comparable tertiary structures. Therefore, neither the addition of glycosites nor their glycosylation compromise the three-dimensional structure of proSap2.

25 Example 6: Immunogenicity of modified proSap2 protein-mannan bioconjugate

FIG. 8 shows preclinical testing of modified proSap2-Fba- β -1,2-mannan bioconjugate ("proSap2-4FM") in rabbit. FIG. 8A shows the modified proSap2-Fba- β -1,2-mannan bioconjugate attributes. FIG. 8B shows a 3D representation of the modified proSap2-Fba- β -1,2-mannan bioconjugate. Structure of a modified proSap2 protein is shown as cartoon. Spheres represent positions of the four introduced glycosites. The conjugated beta-mannan chain is schematically represented, the position and the sequence of the Fba peptide sequence YGKDVKDLFDYAE (SEQ ID NO: 3) is also shown. FIG. 8C shows the rabbit immunization scheme with the modified proSap2-Fba- β -1,2-mannan bioconjugate. As used here, "proSap2-4FM" refers to a modified proSap2-Fba- β -1,2-mannan bioconjugate (modified in that it comprises (i) the inactivating D218N substitution, (ii) the pro-peptide sequence, (iii) 4 glycosites, (iv) the Fba sequence, and (v) a glycan)).

FIG. 9 shows immunogenicity of the purified modified proSap2-Fba- β -1,2-mannan bioconjugate ("proSap2-4FM") in rabbits. FIGS. 9A, 9B and 9C show immunogenicity against Sap2, Fba and β -mannan, respectively, of the modified proSap2-Fba- β -1,2-mannan bioconjugate compared to mSap2 or Sap2-Fba- β -1,2-mannan bioconjugate, ("Sap2-4FM^{unmodif}") in rabbit. ; Control indicates buffer immunized animals, all vaccine groups tested with AS03. New Zealand White (NZW) rabbits immunized three times on a two-week interval. Coating ELISA: FIG. 9A, modified proSap2 not engineered with His-Tag, FIG. 9B, Fba peptide, FIG. 9C, *C.albicans* mannan extract. ProSap2, Fba or mannan-specific serum IgG concentration (arbitrary units, AU) in pre-, post-II, post-III (d0, d28, d42) rabbit sera by treatment group. Lines indicate Geometric Mean Concentration (GMC) $\pm$ 95% confidence interval. ****: $p < 0.0001$, one-way ANOVA. Ref: 36_009, 36_012. As used here, "proSap2-FM" refers to a modified proSap2-Fba- β -1,2-mannan bioconjugate (modified in that it comprises (i) the inactivating D218N substitution, (ii) the pro-peptide sequence, (iii) 4 glycosites, (iv) the Fba sequence, and (v) a glycan); "mSap2" refers to a modified mature Sap2 protein (modified in that it comprises the inactivating D218N substitution, but lacks (i) the pro-peptide sequence, (ii) the fba sequence and (iii) the glycosites; and "Sap2-4FM^{unmodif}" refers to a modified mature Sap2-Fba- β -1,2-mannan bioconjugate (modified in that it comprises: (i) the inactivating D218N substitution, (ii) 4 glycosites, (iii) the Fba sequence, and (iv) a glycan, but lacks the pro-peptide sequence).

Conclusion:

- 1) Modified proSap2-Fba- β -1,2-mannan bioconjugate (proSap2-FM), Sap2-Fba- β -1,2-mannan bioconjugate (Sap2-4FM^{unmodif}) and mSap2 are immunogenic against Sap2. Significant titer increase pre/post-II and pre/post-III shown (FIG. 9A).
- 2) Fba peptide (part of the modified proSap2-Fba- β -1,2-mannan bioconjugate and modified Sap2 protein-glucan bioconjugate) is immunogenic against Fba. Significant titer increase pre/post-II and pre/post-III shown for Sap2-4FM. Sap2-4FM^{unmodif} only showed Significant titer increase pre/post-III (FIG. 9B).
- 3) Modified Sap2-Fba protein- β -1,2-mannan bioconjugate is immunogenic against β -mannan. Significant titer increase pre/post-II and pre/post-III shown (FIG. 9C).

Example 7: Protease activity inhibition assay

- 30 In order to assess whether the antibodies raised in rabbits can bind the active site of Sap2, therefore impairing its activity, an assay was designed in which Sap2 is mixed with a strong inhibitor of its activity (a mono-specific Fab fragment) and subsequently the originated sera are titrated into the solution. If the raised antibodies present in the sera have the same binding epitope of the neutralizing Fab, they should be able to displace them in a concentration-dependent manner.

Firstly, it was proven that the Fab inhibits Sap2 activity. For that, 5 µg/mL anti-Sap2 Fab are complexed with 5 µg/mL proSap2wt in presence of 10 mg/mL BSA at pH 7.5 for 1h at 30°C, shaking at 450 rpm. Activation of proSap2 to mSap2 (cleavage of propeptide) was obtained by pH shift to pH 4 by adding citric acid to 0.1 M, incubating at 30°C for 4 hours, shaking at 450 rpm. Trichloroacetic acid (TCA) was added to 5% (w / w), tubes were incubated for 30 min on ice and centrifuged 10 minutes at 16,000 g to precipitate all the proteins. Peptides resulting from Sap2 digestion of BSA should not precipitate with the rest of the proteins; measurement of the residual absorbance at 280 nm in the supernatant is therefore proportional to the Sap2 protease activity. In Figure 10 A the measured activities in absence of Sap2 (negative control), in absence of the Fab fragment (positive control), and in presence of the Fab fragment (Fab anti-Sap2) were expressed as percentage of the positive control activity as determined via absorbance at 280 nm. The employed Fab fragment was able to completely block the protease activity of Sap2.

Secondly, the ability of the antisera raised in rabbits were tested for their ability to displace the Fab fragment. For this purpose, an ELISA setup was developed where microtiter plates were coated with 1 µg/µL proSap2 in PBS pH 7.4 over night at 4°C, blocked with PBS buffer containing Tween (PBST). Sera from rabbit immunization studies diluted in PBST were titrated into the plates and incubated at 20°C 1 hour, shaking at 450 RPM. Fab contains a His6 tag, and it was pre-incubated with an anti-His-HRP (horseradish peroxidase) secondary antibody in PBST for 1 hour at room temperature, shaking at 450 RPM. The Fab-secondary antibody complex was then added to the wells and plates were washed. TMB (3,3',5,5' tetramethylbenzidine) was added to the plates to detect the presence of the secondary antibody by reacting with the HRP and emitting at 450 nm and therefore to assess whether the Fab fragment binding to proSap2 was compromised by the presence of the sera. In Figure 10 B titration curves are shown where sera from a bicomponent immunization ("Candi5V", where proSap2-Fba-β-1,2-mannan bioconjugate and Als3-Fba-β-1,3-glucan bioconjugate are formulated together) or proSap2-Fba-β-1,2-mannan rabbits immunization "proSap2-4FM") are used. As a control, sera from rabbits with mock immunization was used.

Conclusion:

As shown in FIGS. 10A and 10B, the employed Fab was able to completely block Sap2 protease activity, therefore proving its binding to Sap2 active site. The sera containing antibody raised by proSap2-Fba-β-1,2-mannan (either as monocomponent immunization or as a component of a bicomponent immunization) were able to displace the Fab in a concentration-dependent fashion, proving their ability to bind to Sap2 active site.

Example 8: Quantitative adhesion assay of *C. albicans* hyphae

Fig. 11 shows the capacity of antibodies against proSap2-4FM bioconjugate to inhibit adhesion of *C. albicans* to vaginal epithelial cells. Sera was mixed with *C. albicans* (SC5314), added to epithelial cells (A431) and incubated for 1.5hs. After wash, adhered cells were measured using Concavalin A – Alexa fluor 488. As used here, “proSap2-FM” refers to a modified proSap2-Fba- β -1,2-mannan bioconjugate (modified in that it comprises (i) the inactivating D218N substitution, (ii) the pro-peptide sequence, (iii) 4 glycosites, (iv) the Fba sequence, and (v) a glycan); “mSap2” refers to a modified mature Sap2 protein (modified in that it comprises the inactivating D218N substitution, but lacks (i) the pro-peptide sequence, (ii) the fba sequence and (iii) the glycosites; and “Sap2-4F” refers to a modified proSap2-Fba protein (modified in that it comprises: (i) the inactivating D218N substitution, (ii) 4 glycosites, and (iii) the Fba sequence, but lacks the pro-peptide sequence and is not glycosylated); “control” indicates buffer immunized animals, pre sera of all animals was pooled to create a pre-immunization control. All vaccine groups tested with AS03 adjuvant. NZW rabbits immunized three times on a two-week interval. Graph depicts Mean+SD. ****: $p < 0.0001$, one-way ANOVA.

Conclusion:

As shown in Fig. 11, sera against proSap2-4FM bioconjugate inhibits adhesion of *C. albicans* hyphae to vaginal epithelial cells. Significant reduction of adhesion to epithelial cells is shown compared to Control sera at post-III. Sera against Sap2-4F and mSap2 did not inhibit adhesion to epithelial cells indicating that effect is likely mediated by β -1,2-mannan antibodies.

Example 9: Neutrophile killing assay

Fig. 12 shows the capacity of antibodies against pro Sap2-4FM bioconjugate to mediate neutrophile killing of *C. albicans* hyphae. *C. albicans* hyphae (SC5314) were mixed with sera, added to neutrophiles and incubated. Then, neutrophiles were lysed and Candida CFU were counted after incubation in YPD agar. As used here, “proSap2-FM” refers to a modified proSap2-Fba- β -1,2-mannan bioconjugate (modified in that it comprises (i) the inactivating D218N substitution, (ii) the pro-peptide sequence, (iii) 4 glycosites, (iv) the Fba sequence, and (v) a glycan); “mSap2” refers to a modified mature Sap2 protein (modified in that it comprises the inactivating D218N substitution, but lacks (i) the pro-peptide sequence, (ii) the fba sequence and (iii) the glycosites; and “Sap2-4F” refers to a modified proSap2-Fba protein (modified in that it comprises: (i) the inactivating D218N substitution, (ii) 4 glycosites, and (iii) the Fba sequence, but lacks the pro-peptide sequence and is not glycosylated); “control” indicates buffer immunized animals, pre sera of all animals was pooled to create a pre-immunization control. All vaccine groups tested with AS03 adjuvant. NZW rabbits immunized three times on a two-week interval. Graph depicts Mean+SD. ***: $p < 0.001$; **: $p < 0.01$ against PBS control, one-way ANOVA.

Conclusion:

As shown in Fig. 12, sera against proSap2-4FM bioconjugate and proSap2-F mediate neutrophil killing of *C. albicans* hyphae. Significant increase in killing (reduction of CFUs) was shown vs post-III control sera. Sera against mSap2 did not inhibit adhesion to epithelial cells indicating that effect is likely mediated by β -1,2-mannan and Fba antibodies.

Example 10: Antibody binding to *C. albicans* hyphae

Fig. 13 shows the capacity of antibodies against proSap2-4FM bioconjugate to bind to *C. albicans* using fluorescent microscopy. Cells were observed with differential interference contrast microscopy (DIC) and fluorescent microscopy (secondary antibody Alexa 488). Sera was mixed with *C. albicans* (SC5314) and coloured using Concavalin A – Alexa fluor 488. As used here, “proSap2-FM” refers to a modified proSap2-Fba- β -1,2-mannan bioconjugate (modified in that it comprises (i) the inactivating D218N substitution, (ii) the pro-peptide sequence, (iii) 4 glycosites, (iv) the Fba sequence, and (v) a glycan); “Candi5V” refers to a multicomponent vaccine comprising Sap2-4FM and Als3-Fba- β -1,3-glucan; “control” indicates buffer immunized animals, all vaccine groups tested with AS03. NZW rabbits immunized three times on a two-week interval.

Conclusion:

As shown in Fig. 13, by using laser confocal microscopy it is possible to corroborate the binding of the Sap2-4FM antibodies to the yeast and hyphal segments of the Candida cells.

SEQUENCE LISTINGS

SEQ ID NO: 1 Full-length Wild type Sap2 sequence from *Candida albicans* (with wild type Leader sequence (italicized) and propeptide sequence (underlined); amino acid residue 274 (D) is double underlined)

*MFLKNIFIALAIAL*LDATPTTTKRSAGFVALDFSVVKTPKAFPVTNGOEGKTSKRQAVPVLHNEQVTYAADITV
GSNNQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPSGSSASQDLNTPFKIGYGDGSSSQG
TLYKDTVGF~~GGV~~SIKNQVLADVDSTSIDQILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYS~~LY~~LN~~SP~~DAATG
QIIFFGGVDNAKYSGSLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVD~~SG~~TTIT~~YL~~QQDLADQIIKAFNGKLT
QDSNGNSFYEDCNLSGDVVFNFSKNAKISVPASEFAASLQGGDGPYDKCQLLFDVNDANILGDNFLRSAYIV
YDLDDNEISLAQVKYTSASSISALT

SEQ ID NO: 2 Consensus sequence (artificial sequence)

G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G

SEQ ID NO: 3 Fba sequence

YGKDVKDLFDYAE

5 SEQ ID NO: 4 Consensus sequence (artificial sequence)

D-Q-N-A-T

SEQ ID NO: 5 Consensus sequence (artificial sequence)

D-Q-N-V-T

10

SEQ ID NO: 6 Consensus sequence (artificial sequence)

J-U-B-D/E-X-N-Z-S/T-J-U-B

wherein X is Q (glutamine), Z is A (alanine), each of J and B comprises 1 to 5 Glycine (G) residues and U comprises 1 to 5 serine (S) residues

15

SEQ ID NO: 7 Consensus sequence (artificial sequence)

GSGGGDQ~~NAT~~GSGGG

SEQ ID NO: 8 Consensus sequence (artificial sequence)

20 GSGGGDQ~~NAT~~GSGGGHHHHHHHHHH

SEQ ID NO: 9 Modified Sap2 sequence (intermediate form of Sap2-Fba fusion protein lacking the native leader sequence, comprising the pro-peptide sequence (underlined); comprising 4 glycosites (bolded), comprising an inactivating substitution (double underlined), and further comprising an Fba sequence (dashed undelined)) (artificial sequence)

25

STPTTTKRSAGFVALDFSVVKTPKAFPVTNGQEGKTSKR **GSGG****DQ~~NAT~~** *GSGG*QAVPVLHNEQVYAADITVGSN
NQKLNIVIDTGSSDLWVVP**DQNV**TCQVTYSDQTADFCKQKGTYPGSSASQDLNTPFKIGYDGGSSSQGTLY
KDTVGFGGVSIKNQVLADVDSIDQILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSLEYLN**SPDQ~~NAT~~**ATG
QIIFGGVDNAKYSGSLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTITITYLQQDLADQIIKAFNGKLT
30 QDSNGNSFYEVDCNLSGDVVFNFNFSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGDNFLRSAYIV
YDLDDNEISLAQVKYTSASSISALTYGKDVKDLFDYAE *GSGGG***DQ~~NAT~~** *GSGGG*

SEQ ID NO: 10 Modified Sap2 sequence (mature form of Sap2-Fba fusion protein lacking the native leader sequence or the pro-peptide sequence, comprising 4 glycosites (bolded), comprising an inactivating substitution (double underlined), and further comprising an Fba sequence (dashed undelined)) (artificial sequence)

35

GGG**DQ**NATGGGQAVPVTLHNEQVTYAADITVGSNNQKLNIVIDTGSSDLWVP**DQ**NVTCQVTYSDQTADFC
KQKGTYPDSSASQDLNTPFKIGYDGSSTQGLTKDVTGFGGVSIKNQVLADVSTSIDQGILGVGYKTNEA
GGSYDNVPVTLKKQGVIAKNAYSLLYLN**S**P**DQ**NATGQIIFFGGVDNAKYSGLIALPVTSDRELRLSGSVEVSGKT
INTDNVDVLV**N**SGTTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSKNAKISVPASEFAAS
5 LQGDDGQPYDKCQLLFDVNDANILGDNFLRSAYIVYDLDNEISLAQVKYTSASSISALTYGKDVKDLFDYAQEG
SGGG**DQ**NATGSGGG

SEQ ID NO: 11 *Citrobacter freundii* P079F I WbaD Sequence
MSNNPKKKILVLTTPYPFPVIGGDKLRIYKICQELSKYYDLTLLTIDDIQDLTIPHDEEVFKYVHKIYLPKIKSFFNV
10 LLALPTSTPLQVAYYKSDNFKQKLNELLPAYDATLSHLIRVGHYAKDVNGVNFLEMTDAISLNYKRVREIKTLKSF
KSFIFSLQKRLERYERSIASSFDLTTFISAVDNKFLYPEERTDVIVSGNGVDTNFLQFKNRHIKSQEPVVLIFIGN
MLSLQNMDAVTFFAKKILPLLNEKGNFIFKVIKISEKSRRLISAIPDVIVTGTVDNILETASDGHIGICSMRLGAG
VQNKVLEYMALGMPCVTTTVGFEGIGARDGNDIVIADSPREYVTAIEKLVNDSNYFSSIAINARNFVVSQYSWEM
QLSTFVASVNKLLK

15
SEQ ID NO: 12 *Citrobacter freundii* P079F I WbaC Sequence
MRIVVNNFFYGVLRGIPIYSELVAKLREEGVEVKELRCPKFLYGLPTWIHNFLFIYDQIITPLFGLFHKTKYNIY
PYNLSLIDLFTNKPIVIIHDFISLNKNKNISACYVKFCILSSNRIRNVILISNTTAKIANKLSLFSNARQILLPNTF
FSFKSLSDGVQKEDHGFLLLVSGMGKNKDIDAALYFSPIEYRIPLKILGCGGGRDLLKAKIHGRDEFNTIEILK
20 QIPLEDVVKLYAHCKFVWAHSLAEGYGRALAEKGISGKNILCTRIPAFIEQNCSNVFYNDTESFQEHYFNLIENT
PVVDVCELKEHIKFSEELKKIYEQ

SEQ ID NO: 13 *Citrobacter freundii* P079F I WbaB Sequence
MIVFFTGSYPSPKCGVDYLYKLIANILPHHSNVKVIKNSLLEFIFYAISNRKAIRHVNIQYPTIGYASNYLSAFKPH
25 FVTLIARFLGIRVSITLHEFTSLSSKARFFANLFKIANNIVATSEYEHENLVKFGFNKDRVIVIPIGSNIKESDVKDKT
IDFINFGIISPGKIEDFLYVIEKIRVDYETLKVVLAGYIPDNSEYADKIIAQAKQLNIEFKPNQTEDELSILVGESKR
AILPYSDGISERRGTALAAMINKCVVSYAGNSSEAFNTICMLARDRDELNKLIDALHTDSADDCFIKAYEYAL
ARHWDKVSEKYLRMFYENCSK

30
SEQ ID NO: 14 *Citrobacter freundii* P079F I Wzx Sequence
MCTKFIKKIPSHFVAGSAWGSRFISIFVQFYSIKILLNYLGTNGYALFSLIASFSAWFLVDIGMSTNLQNKISER
KAYGKAYFDLVKRTGCFILVALTLFVILLWIFGPYLSRILLVSFDLSEKDKNIFFISSLLFIGNGIGFFAYKIWYAE
HKGWISNILPAISSICGLLFLILLRTEQFNISHLIIVCLLSFYGPAAFLGVLSFLTTFASSLHYKEKPLSHSIVQDIIRP
SMGFFSFSVMAALVLQVDYIIMSHTLTGKDVIYINVLSKIFGLINFLYAALLQSVWPLCAEAKYKDDKSIYNQIKVK
35 YIGFGAILVFAISLFIYLARDLIILLARDAGISFSFTLIAFFAIYHMIRIWTDTYAMFLLSTGNMRFLYISVPCQALLS
GVLQWYGSVLFGLPGIILGLIMSYILTVSWILPYSFNKNVQGRLCR

SEQ ID NO: 15 *Candida Albicans* Bmt3 Sequence

MKVKVLSLLVPALLVAGAANASDYTPIKVSGYTFKNQVATKNLQCDIVYDQDLQVSAVDLKNKPEDLKFFRD
 KLNELRSLNNIYDLFFQDNEDEVEESILERKWKYFCGSAVWLDKYGVYFMVNRIAYSCKGTRNNPTISVLAGQVF
 5 DKNWIELTGKKFPFSGLEFPTILPHYIDEGKEAEKVLGAEDPRVILHEYTNENGIRIQEPLIAFNALSTVDWKRA
 MHIYRPLHDPHRIIRLSIENYAPREKEKNWAPFIDGNNLNFVYNFPLRILRCNINNGDCQKVSQPDFNDKSHENA
 GKLRGGTNLVEIPSQSLPKHLRSRKYWFGIARSHITDCGCVGELYRPHILISRNKKSDQYELNYVSDLIDFNVNP
 EPWTPGKTTCSDGKSVLIPNSVAFIKDDYMSVTFSEADKTNKLINAKGWLTYYTKMLEFTQERLKDESSDPVLES
 RLLSKCSTFLAQQYCALSKDTMGWDKLSR

10

SEQ ID NO: 16 *Campylobacter* PglB (evolved) sequence

MLKKEYLKNPYLVLFAMIIILAYVFSVFCRFYWWWASEWNEYFHNNQLMIISNDGYAFAEGARDMIAGFHQPN
 DLSYYGSSLSALTYWLYKITPFSFESIILYMSTFLSSLVWIPQILLANEYKRRLMGFVAALLASIANSYNRTMSGYY
 DTDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFIGIYLWWYPSSYTLNVALIGLFLIYTLIFHRKEKIFYIAVILSSL
 15 TLSNIAWFYQSIIIVILFALFALEQKRLNFMIIIGILGSAWLIFLILSGGVDPILYPLKFYIFRSESTNLQGFMYFNV
 VQTIQEVENVDLSEFMRRISGSEIVFLFSLGFWLLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMAIGFGF
 LLSEFKAIMVKYSQLTSCVCIVFATILTAPVFIHIYNYRAPTVFSQNEASLLNQLKNIANREDYVVTWWDYGYF
 VRYYSVDVKTMDVGGKHLGKDNFFPSFALSKEQAAANMARLSVEYTEKSFYAPQNDILKTDILQAMMKDYNQS
 NVDLFLASLSKPDFKIDTPKTRDIYLYMPARMSLIFSTVANFSFINLDTGVLDKPFSTAYPLDVKNGEIYLSNGV
 20 VLSDDFRSFKIGDNVSVNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFNSAYVQMFFLG
 NYDKNLFDLVINSRDAKVFKLKI

SEQ ID NO: 17 Wild type Sap2 sequence (intermediate form of Sap2 or "proSap2") from *Candida albicans* (comprising the wild type propeptide sequence (underlined), but lacking the wild type

25 Leader sequence; the amino acid residue (D) at position 256 is double underlined – this position corresponds to position 274 of SEQ ID NO: 1) (amino acid residues 19-398 of SEQ ID NO: 1)
TPTTTKRSAGFVALDFSVVKTPKAFPVTNGQEGKTSKRQAVPVLHNEQVYAADITVGSNNQKLNVIDTGSS
 DLWVPDVNVDCQVTYSDQTADFCKQKGTYPSPGSSASQDLNTPFKIGYGDGSSSQGTLYKDTVGFGGVSIKNQ
 VLADVSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSLLNSPDAATGQIIFGGVDNAKYSGLI
 30 ALPVTSDRELRLISLGSVEVSGKTINTDNVDVLVDSGGTTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLS
 GDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGDNFLRSAYIVYDLDNEISLAQVKYTS
 ASSISALT

SEQ ID NO:18 Consensus sequence (artificial sequence)

35 K-D/E-X-N-Z-S/T-K
 wherein X is Q (glutamine) and Z is A (alanine)

SEQ ID NO: 19 Consensus sequence (artificial sequence)

K-D-Q-N-A-T

5 SEQ ID NO: 20 Consensus sequence (artificial sequence)

K-D-Q-N-A-S

SEQ ID NO: 21 *E. coli* flagellin (FlgI) signal sequence

MIKFLSALILLLVTTAAQA

10

SEQ ID NO: 22 *E. coli* outer membrane porin A (OmpA) signal sequence

MKKTAIAlAVALAGFATVAQA

SEQ ID NO: 23 *E. coli* maltose binding protein (MalE) signal sequence

15 MKIKTGARILALSALTMMFSASALA

SEQ ID NO: 24 *E. coli* outer membrane porin A (OmpC) signal sequence

MKVKVLSELLVPALLVAGAANA

20 SEQ ID NO: 25 Nucleotide sequence of WbaD of *Citrobacter freundii* P079F I

ATGAGCAATAACCCCAAGAAAAAATTCTTGTTCTAACGCCTCGCTATCCATTTCTGTAATCGGTGGTGATA
AACTGAGAATTTATAAAATATGTCAGGAACCTTTCAAATACTATGATCTTACATTATTAACTTTAATTGATGAT
ATTCAGGATTTAACCATACCTCATGATGAGGAAGTTTTAAATACGTACATAAGATATATCTCCCTAAAATAA
AGTCTTTTTTAAATGTACTTTTGGCATTACCTACTTCTACGCCATTGCAGGTTGCATATTATAAATCTGATAA
25 TTTTAAGCAAAAACCTGAATGAATTGTTGCCTGCCTATGATGCAACACTTTCTCATTTGATCAGAGTTGGGCAT
TATGCCAAGGATGTGAATGGTGTGAATTTCTTGAAATGACAGATGCCATATCTCTTAACACAAAAGAGTTA
GGGAAATTAACACTGAAAAGTTTTAAATCCTTCATTTTTCTTTAGAGCAAAAAGATTGGAGCGCTATGA
ACGTTCAATTGCAAGTTCTTTCGATTTAACGACATTTATTTCTGCGGTGGATAAGAATTTTTATATCCAGAG
GAAAGAACTGATGAATTGTATCAGGGAATGGTGTAGATACAAATTTCTTCAGTTTAAAAATCGCCATATTA
30 AATCACAAGAACCTGTTGTTTTGATTTTTATCGGGAATATGCTCTCATTGCAAAATATGGATGCTGTCACATT
CTTTGCGAAAAAATACTACCACTTCTTAACGAAAAAGGTAATTTTCATATTTAAAGTTATCGGTAAAATTTCT
GAGAAGAGTAGAAGGATTCTGTCCGCAATCCCTGATGTGATAGTCACTGGTACTGTTGATAATATTTAGAA
ACTGCATCTGATGGGCACATTGGTATTTGCTCAATGAGATTGGGTGCTGGAGTACAGAATAAAGTATTAGAA
TATATGGCCTTGGGTATGCCTTGTAACACTACGACTGTGGGGTTTGAAGGCATTGGGGCTAGGGATGGTAA
35 CGACATAGTTATTGCTGACTCACCCCAGAAATATGTAACCTGCAATTGAAAAATTAGTCAATGACAGCAATTAT

TTTTCTTCTATAGCAATCAATGCAAGGAATTTTGTGTGTCCCAACTCTTGGGAAATGCAGTTGTCGACAT
TTGTTGCTTCTGTAAATAAGCTCCTAAAATAA

SEQ ID NO: 26 Nucleotide sequence of WbaC of *Citrobacter freundii* P079F I

- 5 ATGAGAATTGTAGTAAATAATTTCTTTTATGGTGTGCTAAACGTGGAATCCCTATTTATACTTCTGAGTTAG
TCGCTAAATTAAGAGAGGAAGGAGTTGAGGTTAAAGAACTAAGATGTCCTAAGTTTCTTTACGGCTTACCTA
CTTGGATTCAATTTTCTTTTATTATTTACGATCAAATTATTACTCCTCTATTTGGTCTTTTTCATAAAACA
AAATATAACATTTATCCATACAACAGTTTATCCTTAATAGACTTGTTCACTAATAAGCCTATAGTTATTATTCA
TGATTTTATTAGTCTAAATAAGAATAAAAAAATATATCTGCATGTTATGTAAATTTTGTATTCTTTCTAGTT
10 CAAATAGAATTAGAAATGTAATTTTAACTCTCAACACGACAGCGAAAATAGCGAATAAATTATCATTATTTTC
AAATGCAAGGCAAATATTATTGCCAAATACTTTCTTTTCATTTAAATCATTGAGTGATGGTGTCCAAAAGGAA
GATCATGGTTTTTTATTATTAGTTTCAGGAATGGGGAAAAATAAGGACATTGATGCGGCTTTAGAAGTTTACT
TTTCAATTCCTATTGAATATCGAATTCGCTAAAAATTTTAGGATGCGGGGGGGGAGGGATTTACTGAAAG
CAAAAATTCATGGACGAGATGAATTCATACTATTGAAATTTTAAACAAATCCCATTAGAGGATGTGGTTAA
15 ACTATACGCGCATTGCAAGTTTGTGTTGGGCCACTCTCTTGCTGAAGTTACGGAAGGGCGCTGGCGGAAGG
TAAGATCTCAGGGAAAAATATTTGTGCACTAGAATTCCTGCATTTATAGAGCAAAATGTTCTAATGTGTTT
TATTATAACGATACTGAAAGCTTCCAAGAGCATTATTTTAAATTTGATTGAGAATACACCGGTTGTCGACGTGT
GTGAGTTAAAAGAACATATAAAATTCAGCGAAGAGTTAAAGAAAATTTATGAGCAATAA

- 20 SEQ ID NO: 27 Nucleotide sequence of WbaB of *Citrobacter freundii* P079F I

- ATGATAGTTTTTTTTACAGGCTCATATCCTCCGAGTAAGTGTGGTGTGGTGATTATTTGTATAAATTAATAG
CTAATATTTTACCACACCATTCAAATGTTAAAGTTATAAAAACTCATTGCTTGAATTTATTTTTATGCCATT
TCAAATAGGAAGGCCATAAGACATGTAAATATACAATATCCGACAATTGGCTATGCTAGTAATTATTTAAGTG
CATTAAACCTCATTTTGTGACACTTATAGCTAGATTTTGGGGATTAGAGTTTCAATAACTCTTCATGAGTT
25 TACTAGCCTTTCTAGTAAGGCGCGGTTTTTGTCTAATCTTTTTAAGATTGCTAATAATATCGTTGCTACATCT
GAGTATGAACATGAGAATCTAGTTAAATTTGGTTTTAATAAGATAGGGTAATAGTTATTCCAATTGGATCGA
ATATAAAAGAGTCCGATGTGAAAGACAAAATATAGACTTTATTAACCTTTGGAATTATACACCAGGAAAAGG
TATTGAGGATTTCTGTATGTTATTGAAAAAATCAGAGTAGATTATGAACTTTGAAAGTTGTTTTAGCTGGC
TATATACCTGACAATAGTGAGTACGCTGATAAATCATTGCGCAAGCCAAACAGCTTAATATTGAATTCAAAC
30 CTAACCAAATGAAGATGAAGTGTCTATCTTAGTTGGGGAATCCAAAAGAGCAATATTGCCTTATAGTGATG
GTATTTAGAGAGAGAGGCACTGCGCTTGACGCTATGATTAACAAATGTGTTGTATATTCGTACGCTGGCA
ATAGTTCAGAAGCATTTAATACGATATGTATGCTAGCGCGTGACAGGGATGAGTTATATAATAAACTCATTGA
TGCTTTACACACTGACTCTGCAGATGATTGCTTTATTGCTAAAGCATATGAATATGCTTTAGCACGCCATTGG
GACAAAGTTTCAGAAAAATACTTAAGGATGTTTTATGAGAATTGTAGTAAATAA

35

SEQ ID NO: 28 Nucleotide sequence of Wzx of *Citrobacter freundii* P079F I

ATGTGCACTAAATTTATAAAAAAATCCCTAGTCATTTTGTGTAGCAGGTAGTGCGTGGGGGAGTCGATTT
ATATCTATATTTGTACAATTCTATAGTATAAAGATATTATTAAGTATTTGGGTACAAACGGGTATGCTCTTT
TCTCATTAATTGCGAGTTTCTCTGCATGGTTTCTCTTGGTTGATTTGGTATGTCAACCAATTTGCAAAATAA
5 AATATCGGAGCGTAAAGCCTATGGAAAGGCATATTTTGATTTGGTTAAAAGAACTGGTTGTTTTTAATTGTA
GCTTTGACTCTATTTGTAATTTTGCTTTGGATATTTGGTCCTTATTTATCGAGGATATTATTGGTGTCTTTG
ATTTTTATCTGAAAAAGACAAAATAATATTTCTTTATTTTCTCTGTTATTCATAGGTAACGGCATTGGT
TTTTTGCATACAAAATTTGGTATGCAGAACATAAAGGTTGGATTTCTAATATACTACCCGCTATATCTTCCA
TATGTGGGTTGTTGTTTTAATCTTACTAAGAACTGAACAATTTAATATAAGTCATTTAATTATTGTATGCCT
10 CTTATCATTCTATGGTCCCGCAGCTTTTCTGGGTGTACTTAGTTTTTAAACAACATTTGCTAGTTCGCTACAT
TATAAGAAAAACCTTTATCTCACTCCATCGTGCAGGATATTATTAGGCCTTCTATGGGTTTTTTTTCATTTT
CTGTAATGGCTGCATTGGTACTACAAGTAGATTATATTATTATGTCCCATACACTAACAGGAAAAGACATTGT
AATATATAATGTGTTAAGTAAAATATTCGGGTAAATCAATTTCTATATGCAGCGTTGCTACAATCAGTTTGG
CCATTATGTGCAGAAGCAAAATATAAAGATGATAAAAGTATTTACAATCAGATAAAAGTAAAGTATATCGGCT
15 TTGGTGCAATATTGGTATTCGCAATTTCTTTATTTATCTATCTTGCTAGGGATTTGATTATATTATTAGC
ACGTGATGCAGGTATTTTCTTTTCTTCTACTCTCATTGCTTTTTTGAATCTATCACATGATAAGAATATGG
ACAGATACGTATGCTATGTTTTTACTTAGTACGGGGAATATGAGGTTCTGTATATATCTGTGCCATGCCAA
GCTTTGTTGAGTGGAGTGTGCAATGGTATGGTTCTGTTTTATTTGGACTTCCAGGGATTATATTAGGCTTA
ATTATGCTTATATATTAACAGTCAGCTGGATATTACCATATTCATTCAATAAAAATGTTCAAGGTAGATTGT
20 GTCGATAA

SEQ ID NO: 29 Nucleotide sequence of Bmt3 of *C. albicans*

ATGAAAGTGAAAGTGCTGAGCCTGCTGGTTCCGGCACTGCTGGTTGCAGGTGCCGCAATGCTAGCGACTAC
ACCCGATTAAAGTGAGCGGTTATACCTTCAAAAACAGGTTGCGACCAAGAACCTGCAATGCGATAGCATC
25 GTGTACGACCAGGATCTGGACCTGCAGGTGTCTCAAGCGGTTGATCTGAACAAGCCGAGGACCTGAAATTC
TTTCGTGATAAGCTGAACGAACTGCGTAGCTGAACAACATTTACGACCTGTTCTTTCAAGATAACGAGGAC
GAAGTGGAGGAAAGCATCCTGGAGCGTAAGTGGTATAAATCTGCGGTAGCGCGGTTTGGCTGGATAAATA
CGGCGTGTATTTTATGGTTAACCGTATCGCGTATAGCAAGAAAGGTACCCGTAACAACCCGACCATTAGCGT
GCTGGCGGGCCAGGTTTTCGACAAGAACTGGATCGAGCTGACCGGTAAGAAATCCCGTTTAGCGGCCTGG
30 AGTTTCCGACCATCCTGCCGCACTACATTGACGAGGGTAAAGAGGCGGAAAAAGTGATCCTGGGCGCGGAA
GATCCGCGTGTTATTTCTGCACGAGTATACCAACGAAACGGTATCCGTATTCAAGAGCCGCTGATTGCGTTC
AACGCGCTGAGCACCGAAGTTGACTGGAAACGTGCGATGCACATTTACCGTCCGCTGCACGATCCGCACCGT
ATCATTCGTCTGAGCATCGAAAATATGCGCCGCGTGAGAAGGAGAAGAACTGGGCGCGGTTTCATCGACGGC
AACAACTGAACTTCGTGTACAATTTCCGCTGCGTATCCTGCGTTGCAACATTAAACAGGTGATTGCCAGA
35 AAGTTAGCGGCCCGGATTTTAAACGACAAAAGCCACGAGAACGCGGGCAAGCTGCGTGGTGGCACCAACCTG
GTGGAAATCCCGAGCCAAAGCCTGCCGAAACACCTGCGTAGCCGTAAGTACTGGTTCGGCATCGCGCTAGC

CACATTACCGACTGCGGTTGCGTTGGCGAGCTGTATCGTCCGCACCTGATCCTGATTAGCCGTAACAAGAAA
AGCGATCAGTACGAGCTGAACTATGTGAGCGATCTGATCGACTTTAACGTTAACCCGGAACCGTGACCCCG
GGTAAACACCTGCAGCGACGGCAAGAGCGTGCTGATCCCGAACAGCGTTGCGTTCATTAAGGACGATTAC
ATGAGCGTGACCTTTAGCGAAGCGGATAAGACCAACAACTGATCAACGCGAAAGGTTGGCTGACCTATATT
5 ACCAAGATGCTGGAGTTCACCCAAGAACGTCTGAAAGATGAGAGCAGCGACCCGGTTCTGGAAAGCCGTCTG
CTGAGCAAGTGCAGCACCTTTCTGGCGCAGCAATACTGCGCGCTGAGCAAAGATACCATGGGCTGGGACAAG
CTGAGCCGTTAA

SEQ ID NO: 30 Nucleotide sequence of *Campylobacter* PglB (evolved)

10 ATGCTGAAGAAGGAATATCTGAAGAACCCGATCTGGTGCTGTTTGCATGATTATCCTGGCGTATGTTTTT
AGTGTGTTTTGTCGTTTCTACTGGGTGTGGTGGGCCAGTGAATGGAACGAATATTTCCACAACAACAGCTG
ATGATCATCTCCAATGATGGCTATGCCTTCGCAGAAGGTGCCCGTGACATGATTGCAGGCTTTCATCAGCCG
AACGATCTGAGTTATTACGGTAGCTCTCTGTCCGCGTGACCTATTGGCTGTACAAAATCACGCCGTTTAGTT
TCGAATCCATTATCCTGTACATGAGTACCTTCCTGAGTTCCTGGTGGTTATTCCGCAGATCCTGCTGGCCAA
15 TGAATATAAACGTCGGCTGATGGGCTTTGTTGCGGCCCTGCTGGCTAGTATTGCGAACTCCTATTACAATCG
CACCATGAGTGGTTATTACGATACGGACATGCTGGTCATTGTGCTGCCGATGTTTCATCCTGTTTTTCATGGT
GCGTATGATTCTGAAAAAGGATTTCTTTAGCCTGATCGCCCTGCCGCTGTTTATTGGCATCTATCTGTGGTG
GTACCCGTCATCGTATACCTGAACGTTGCACTGATTGGTCTGTTTCTGATTTACACGCTGATCTTCCATCGC
AAGGAAAAGATCTTTTATATCGCGTTATCTTAAGCTCTCTGACCCTGAGCAACATTGCTTGGTTTTATCAGT
20 CTGCGATTATCGTCATCCTGTTTGCCCTGTTGCACTGGAACAAAAACGTCTGAATTTTCATGATTATCGGCAT
TCTGGGTAGTGCCTGGCTGATCTTTCTGATTCTGTCCGCGGTGTTGATCCGATTCTGTACCCGCTGAAATT
TTATATCTTCCGCTCAGATGAGTCGACCAACCTGACCCAAGGCTTCATGTACTTCAACGTTGTGCGACGATC
CAAGAAGTGGAAAATGTTGATCTGAGCGAATTTATGCGTCGCATTAGTGGCTCCGAAATCGTTTTTCTGTTT
TCACTGTTAGTTTTCGTCTGGCTGCTGCGTAAACACAAGTCGATGATTATGGCCCTGCCGATTCTGGTGCTG
25 GGTTTTCTGGCACTGAAAGGCGGTCTCCGCTTTACCATTTACAGCGTTCCGGTCATGGCCCTGGGCTTTGGT
TTCCTGCTGTCTGAATTTAAGGCAATCATGGTTAAAAAGTACTCACAGCTGACCTCGTGCCTCTGCATTGTGT
TCGCCACCATCCTGACGCTGGCACCAGGTGTTTCATCCATATCTACAACACAGGGCTCCGACGGTGTAGCC
AGAACGAAGCGTCGCTGCTGAATCAACTGAAGAACATTGCCAATCGTGAAGATTATGTCGTGACCTGGTGGG
ACTATGGCTACCCGGTGCGCTATTACAGCGATGTTAAACGATGGTCGACGGCGGTAAACACCTGGGCAAGG
30 ATAACCTTTTCCCGAGCTTTGCTCTGTCTAAAGATGAACAGGCAGCTGCGAATATGGCGCGCTGTGAGTCG
AATACACCGAAAAGTCGTTTTATGCCCGCAGAATGATATTCTGAAAACGGACATCCTGAGGCAATGATGA
AGGATTATAACCAAAGCAATGTTGACCTGTTCTGGCATCACTGTCGAAACCGGATTTTAAGATTGACACCCC
GAAAACGCGTGATATCTATCTGTACATGCCGGCTCGCATGAGTCTGATTTTTAGCACCGTCGCGAACTTTTCT
TTCATCAACCTGGATACGGGCGTGCTGGACAAACCGTTTACCTTCTCAACGCGTACCCGCTGGATGTGAAG
35 AACGGCGAAATTTATCTGTGCAATGGTGTGCTGAGCGATGACTTTCTGTTCTTTCAAATCGGCGATAAC
GTTGTGAGCGTGAACAGCATCGTTGAAATTAATAGCATCAAACAGGTGAATACAAGATTACCCCGATCGAT

GACAAGGCTCAATTCTACATTTTCTACCTGAAGGACTCCGCTATTCCGTATGCGCAGTTCATCCTGATGGATA
AAACCATGTTTAACTCTGCGTACGTGCAAATGTTTTCTGGGTAACTACGATAAGAACCTGTTTGACCTGGT
CATTAAATCTCGCGATGCTAAGGTGTTTAACTGAAGATCTAA

5 SEQ ID NO: 31 *E. coli* K12 W3110 manB sequence

MKKLTCKAYDIRGKLGEELNEDIAWRIGRAYGEFLKPKTIVLGGDVRLTSETLKLALAKGLQDAGVDVLDIGMS
GTEEIYFATFHLGVDGGIEVTASHNPMDDYNGMKLVREGARPISGDTGLRDVQRLAEANDFPVDET KRGRYQQI
NLRDAYVDHLFGYINVKNLTPLKLVINSNGAAGPVVDAIEARFKALGAPVELIKVHNTPDGNFPNGIPNLLPEC
RDDTRNAVIKHGADMGIADFDFRCLFDEKGQFIEGYYIVGLLAEAFLEKNPGAKIIHDPRLSWNTVDVVTAA
10 GGTPVMSKTGHAFIKERMKEDAIYGGEMSAHHYFRDFAYCDSGMIPWLLVAELVCLKDKTLGELVRDRMAAFP
ASGEINSKLAQPVEAINRVEQHFSSREALAVDRDTGISMFTFADWRFNLRTSNTEPVVRLNVESRGDVPLMEARTR
TLLTLLNE

SEQ ID NO: 32 *E. coli* K12 W3110 manC sequence

15 MAQSKLYPVMMAGGSGSRLWPLSRVLYPKQFLCLKGLDMLQTTICRLNGVECESPVVICNEQHRFIVAEQLRQL
NKL TENIILEPAGRNTAPAIALAALAAKRHSPESDPLMLVLAADHVIADDAFRAAVRNAMPYAEAGKLVTFGIVP
DLPETGYGYIRRGESVAGEQDMVAFEVAQFVEKPNLETAQAYVASGEYYWNSGMFLFRAGRYLEELKKYRPDIL
DACEKAMSAVDPDLNFIRVDEEAFACPEESVDYAVMERTADAVVVPMDAGWSDVGSWSSLWEISAHTAEGNV
CHGDVINHK TENSYYAESGLVTTVGVDLVVVQTKDAVLIADRNAVQDVKKVVEQIKADGRHEHRVHREVYR
20 PWGKYDSIDAGDRYQVKRITVKPGEGLSVQMHHHRAEHWWVVAGTAKVTIDGDIKLLGENESIYIPLGATHCLE
NPGKIPLDLIEVRSGSYLEEDDVVRFADRYGRV

SEQ ID NO: 33 Nucleotide sequence of manB from *E. coli* K12 W3110

ATGAAAAAATTAACCTGCTTTAAAGCCTATGATATTCGCGGAAATTAGGCGAAGAACTGAATGAAGATATC
25 GCCTGGCGCATTGGTCGCGCTATGGCGAATTTCTCAAACCGAAAACCATTTGTGTTAGGCGGTGATGTCCGC
CTCACCAGCGAAACCTTAAACTGGCGCTGGCGAAAGTTTACAGGATGCGGGCGTTGACGTGCTGGATATT
GGTATGTCCGGCACCGAAGAGATCTATTCGCCACGTTCCATCTCGGCGTGGATGGCGGCATTGAAGTTACC
GCCAGCCATAATCCGATGGATTATAACGGCATGAAGCTGGTTCGCGAGGGGGCTCGCCCGATCAGCGGAGA
TACCGGACTGCGCGACGTCCAGCGTCTGGCTGAAGCCAACGACTTTCCTCCCGTCGATGAAACCAAACGCGG
30 TCGCTATCAGCAAATCAACCTGCGTGACGCTTACGTTGATCACCTGTTGCGTTATATCAATGTCAAAAACCTC
ACGCCGCTCAAGCTGGTGATCAACTCCGGGAACGGCGCAGCGGGTCCGGTGGTGACGCCATTGAAGCCCG
CTTTAAAGCCCTCGCGCGCCCGTGAATTAATCAAAGTGACAAACACGCCGACGGCAATTTCCCAACGG
TATTCCTAACCCACTACTGCCGGAATGCCGCGACGACCCGCAATGCGGT CATCAAACACGGCGCGGATAT
GGGCATTGCTTTTGATGGCGATTTTGACCGTGTTCCTGTTTGACGAAAAAGGCAGTTTATTGAGGGCTA
35 CTACATTGTGCGCCTGTTGGCAGAAGCATTCCTCGAAAAAATCCCGGCGCAAGATCATCCACGATCCACG
TCTCTCCTGGAACACCGTTGATGTGGTGACTGCCGCGAGTGGCACGCCGTAATGTGAAAACCGGACACGC

CTTTATTAAAGAACGTATGCGCAAGGAAGACGCCATCTATGGTGGCGAAATGAGCGCCACCATTACTTCCG
TGATTTTCGCTTACTGCGACAGCGGCATGATCCCGTGGCTGCTGGTCGCCGAACCTGGTGTGCCTGAAAGATAA
AACGCTGGGCGAACTGGTACGCGACCGGATGGCGGCGTTTCCGGCAAGCGGTGAGATCAACAGCAAACCTGG
CGCAACCCGTTGAGGCGATTAACCGCGTGGAACAGCATTTTAGCCGTGAGGCGCTGGCGGTGGATCGCACC
5 GATGGCATCAGCATGACCTTTGCCGACTGGCGCTTTAACCTGCGCACCTCCAATACCGAACCGGTGGTGC GC
CTGAATGTGGAATCGCGCGGTGATGTGCCGCTGATGGAAGCGCGAACGCGAACTCTGCTGACGTTGCTGAA
CGAGTAA

SEQ ID NO: 34 Nucleotide sequence of manC from *E. coli* K12 W3110

10 ATGGCGCAGTCGAACTCTATCCAGTTGTGATGGCAGGTGGCTCCGGTAGCCGCTTATGGCCGCTTTCCCGC
GTACTTTATCCCAAGCAGTTTTATGCCTGAAAGGCGATCTCACCATGCTGCAAACCACCATCTGCCGCTGA
ACGGCGTGGAGTGCGAAAGCCCGGTGGTGAATTTGCAATGAGCAGCACCGCTTTATTGTGCGGGAACAGCTG
CGTCAACTGAACAACTTACCGAGAACATTATTCTGAACCGGCGAGGCGAAACACGGCACCTGCCATTGCG
CTGGCGGCGCTGGCGGCAAAACGTATAGCCCGGAGAGCGACCCGTTAATGCTGGTATTGGCGGCGGATCA
15 TGTGATTGCCGATGAAGACGCGTTCGCTGCCGCGTGCCTAATGCCATGCCATATGCCGAAGCGGCAAGCT
GGTGACCTTCGGCATTGTGCCGATCTACCAGAAACCGGTTATGGCTATATTCTGCGCGGTGAAGTGTCTGC
GGGTGAGCAGGATATGGTGGCCTTTGAAGTGGCGCAGTTTGTGAAAAACCGAATCTGGAACCGCTCAGG
CCTATGTGGCAAGCGGCGAATATTACTGGAACAGCGGTATGTTCTGTTCCGCGCCGACGCTATCTCGAAG
AACTGAAAAAATATCGCCCGGATATCCTCGATGCCTGTGAAAAAGCGATGAGCGCCGTCGATCCGGATCTCA
20 ATTTTATTGCGTGGATGAAGAAGCGTTTCTGCGCTGCCCGGAAGAGTCGGTGGATTACGCGGTGATGGAAC
GTACGGCAGATGCTGTTGTGGTGCCGATGGATGCGGGCTGGAGCGATGTTGGCTCCTGGTCTTCATTATGG
GAGATCAGCGCCACACCGCCGAGGGCAACGTTTGCCACGGCGATGTGATTAATCACAAAACCTGAAAACAGC
TATGTGTATGCTGAATCTGGCCTGGTCACCAACCGTCGGGGTGAAAGATCTGGTAGTGGTGCAGACCAAAGAT
25 TCGCCATGAGCATCGGGTGCATCGCGAAGTGTATCGTCCGTGGGGCAAATATGACTCTATCGACGCGGGCG
ACCGCTACCAGGTGAAACGCATCACCGTGAAACCGGGCGAGGGCTTGTCCGTACAGATGCACCATCACCGCG
CGGAACACTGGGTGGTTGTGCGGGGAACGGCAAAAGTCACCATTTGATGGTGATATCAAACCTGTTGGTGAAA
ACGAGTCCATTTATATTCCGCTGGGGGCGACGCATTGCCTGGAACCCGGGGAAAAATTCCGCTCGATTTAA
TTGAAGTGCCTCCGGCTCTTATCTCGAAGAGGATGATGTGGTGCCTTCGCGGATCGCTACGGACGGGTG
30 TAA

SEQ ID NO: 35 Nucleotide sequence of wild type Sap2 from *C. albicans*

ATGTTTTTAAAGAATATTTTCATTGCTCTTGCTATTGCTTTATTAGTCGATGCTACTCCAACAACAACCAAAA
GATCAGCTGGTTTCGTTGCTTTAGATTTTCAGTGTTGTGAAAACTCCTAAAGCATTCCAGTTACTAATGGTCA
35 AGAAGGTAAAACTTCCAAAAGACAAGCTGTCCAGTGACTTTACACAATGAACAAGTCACTTATGCTGCTGAT
ATTACCGTTGGATCCAATAATCAAAAACCTTAATGTTATTGTTGATACTGGATCATCTGATTTATGGGTTCTG

- ATGTTAATGTTGATTGTCAAGTCACTTATAGTGATCAAAGTGCAGATTTCTGTAAACAAAAGGGGACATATGA
TCCAAGTGGTTCATCAGCTTCACAAGATTTGAATACTCCATTCAAATTTGGTTATGGTGATGGATCTTCATCT
CAAGGTACTTTATATAAGGATACCGTTGGATTTGGTGGTGTTCGATTAAAAATCAAGTTTTAGCTGATGTT
GATTCTACTTCAATTGATCAAGGTATTTAGGGGTTGGTTATAAAACCAATGAAGCCGGTGGTAGTTATGAT
5 AATGTCCTGTCACTTTAAAAAACAAGGAGTCATTGCTAAGAATGCTTATTCACTTTATCTTAATTCTCCAG
ATGCTGCCACGGGACAAATAATTTTCGGTGGGGTTGATAATGCTAAATATAGTGGTTCATTAATTGCATTAC
CAGTTACTTCTGATCGTGAATTAAGAATTAGTTTGGGTTGAGTTGAAGTTTCTGGTAAAACCATCAATACTGA
TAATGTCGATGTTCTTGTGGATTGAGTACCACCATTAATTTGCAACAAGATCTTGCTGATCAAATCATT
AAAGCTTTCAATGGTAAATTAAGTCAAGATTCAATGGTAATTCATTCTATGAAGTTGATTGTAATTTGTCAG
10 GGGATGTTGTATTCAATTTTAGTAAAAATGCTAAAATTTCCGTTCCAGCTTCCGAATTTGCTGCTTCTTTACA
AGGTGATGATGGTCAACCATATGATAATGTCAATTAATTTTCGATGTTAATGATGCTAACATTTCTTGGTGAT
AACTTTTTGAGATCAGCTTATATTGTTTATGATTGGATGATAATGAAATTTCTTTGGCTCAAGTCAAATATA
CTTCTGCTTCCAGTATTTCTGCCTTGACCTAA
- 15 SEQ ID NO: 36 Nucleotide sequence of modified Sap2 (intermediate form of Sap2-Fba fusion protein
without the native leader sequence) comprising the pro-peptide sequence; comprising 4 glycosites
and inactivating substitution)
AGCACCCCGACCAACCAAGCGTAGCGCGGGCTTTGTTGCGCTGGACTTTAGCGTTGTGAAGACCCCGAAG
GCGTTTCCGGTTACCAATGGCCAGGAAGGTAAAACAGCAAACGTGGTGGCGGCGATCAGAACGCGACCGG
20 CGGCGGCCAAGCGGTGCCGTTACCTGCACAACGAACAGGTTACCTACGCGGCGGATACACCGTGGGCA
GCAACAACCAAAAGCTGAACGTGATTGTTGACACCGGTAGCAGCGATCTGTGGTTCGGGACCAAGCTTA
CCTGCCAGGTGACCTACAGCGACCAAAACCGCGGATTTCTGCAAGCAGAAAGGCACCTATGACCCGAGCGGTA
GCAGCGGAGCCAAGATCTGAACACCCCGTTCAAATTTGGTTATGGTGATGGTAGCAGCAGCCAGGGTACCC
TGTATAAGGATACCGTTGGTTTTGGTGGCGTGAGCATCAAAAACAGGTTCTGGCGGACGTGGATAGCACCA
25 GCATCGACCAAGGCATTCTGGGCGTTGGTTACAAGACCAACGAGGCGGGTGGCAGCTATGATAACGTGCCG
GTTACCTGAAGAAACAGGGTGTGATCGCGAAAAACGCGTACAGCCTGTATCTGAACAGCCCGGATCAGAAC
GCGACCGGTCAAATCATTTTCGGTGGCGTTGATAACGCGAAGTACAGCGGTAGCCTGATTGCGCTGCCGGT
GACCAGCGACCGTGAGCTGCGTATTAGCCTGGGCAGCGTGGAAGTTAGCGGTAAAACCATCAACACCGACAA
CGTGGATGTTCTGGTGAATAGCGGCACCACTTACCTACCTGCAGCAAGACCTGGCGGATCAGATCATTA
30 GGCGTTCAACGGCAAAGTACCAAGACAGCAACGGAACAGCTTTTATGAGGTTGACTGCAACCTGAGCGG
TGATGTGGTTTTCACTTTAGCAAGAACGCGAAATCAGCGTGCCGGCGAGCGAATTTGCGGCGAGCCTGCA
GGGCGACGATGGTCAACCGTATGACAAGTGCCAGCTGCTGTTGACGTTAACGATGCGAACATCTGGGTGA
CAACTTTCTGCGTAGCGGTACATTGTGTATGATCTGGACGATAACGAAATTAGCCTGGCGCAAGTTAAATA
CACCAGCGGAGCAGCATTAGCGCGTGACCTATGGCAAAGATGTGAAAGATCTGTTTGATTATGCGCAGGA
35 AGGCTCTGGCGGCGGCGATCAGAACGCGACCGGCTCCGGCGGCGGCCACCATCACCATCACCATCACCATCA
CCATTAA

SEQ ID NO: 43 SipA sequence

MKMNKKVLLTSTMAASLLSVASVQAS

5 SEQ ID NO: 44 proSap2 N Sequence

MKIKTGARILALSALTMMFSASALASGGGGGQDNATKGGGGTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQ
EGKTSKRQAVPVT LHNEQVYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQV TYSDQTADFCKQKGT
DPSGSSASQDLNTPFKIGYGDGSSSQGTLYKDTVGGGVS IKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDN
VPVTLKKQGVIAKNAYSLYLNSPDAATGQIIFGGVDNAKYSGLIALPVTS DRELRLSLGSVEVSGKTINTDNVDV
10 LVNSGTTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSKNAKISVPASEFAASLQGDDGQP
YDKCQLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHH

SEQ ID NO: 45 proSap2 N3C Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRLEVLFGPGG
15 DQDNATKGGGQAVPVT LHNEQVYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQV TYSDQTADFCKQKGT
TYDPSGSSASQDLNTPFKIGYGDGSSSQGTLYKDTVGGGVS IKNQVLADV DSTSIDQGILGVGYKTNEAGGSY
DNVPVTLKKQGVIAKNAYSLYLNSPDAATGQIIFGGVDNAKYSGLIALPVTS DRELRLSLGSVEVSGKTINTDNV
DVLVNSGTTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSKNAKISVPASEFAASLQGDDG
QPYDKCQLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHH
20

SEQ ID NO: 46 proSap2 C Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVT LHNE
QVYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQV TYSDQTADFCKQKGTYPDSGSSASQDLNTPFKI
YGDGSSSQGTLYKDTVGGGVS IKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
25 YLNSPDAATGQIIFGGVDNAKYSGLIALPVTS DRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
IIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILG
NFLRSAYIVYDLDDNEISLAQVKYTSASSISALTGGGGGKDQDNATGGGGGHHHHHHHHH

SEQ ID NO: 47 proSap2 Mut1a Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVT LHND
30 QNATYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQV TYSDQTADFCKQKGTYPDSGSSASQDLNTPFKI
GYGDGSSSQGTLYKDTVGGGVS IKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYS
LYLNSPDAATGQIIFGGVDNAKYSGLIALPVTS DRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLAD
QIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILG
35 DNFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHH

SEQ ID NO: 48 proSap2 Mut1b Sequence

MKIKTGARILALSALTTMMFSASALASTPTTTTKRSAGFVALDFSVVKTPKAFPVTNGQEGKTSKRQAVPVTLHNK
DQNATGYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPGSSASQDLNTP
5 FKIGYGDGSSSQGTLYKDTVGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKN
AYSLYLNSPDAATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQD
LADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDA
NILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

10 SEQ ID NO: 49 proSap2 Mut2a Sequence

MKIKTGARILALSALTTMMFSASALASTPTTTTKRSAGFVALDFSVVKTPKAFPVTNGQEGKTSKRQAVPVTLHNE
QVTYAADITVDQNA TQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPGSSASQDLNTPFKI
GYGDGSSSQGTLYKDTVGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYS
LYLNSPDAATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLAD
15 QIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILG
DNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 50 proSap2 Mut2b Sequence

MKIKTGARILALSALTTMMFSASALASTPTTTTKRSAGFVALDFSVVKTPKAFPVTNGQEGKTSKRQAVPVTLHNE
20 QVTYAADITVKDQNA TQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPGSSASQDLNTPF
KIGYGDGSSSQGTLYKDTVGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNA
YSLYLNSPDAATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDL
ADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDAN
ILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

25

SEQ ID NO: 51 proSap2 Mut2c Sequence

MKIKTGARILALSALTTMMFSASALASTPTTTTKRSAGFVALDFSVVKTPKAFPVTNGQEGKTSKRQAVPVTLHNE
QVTYAADITVGSNKDQNA TQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPGSSASQDLN
TPFKIGYGDGSSSQGTLYKDTVGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAK
30 NAYSLYLNSPDAATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQ
DLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSNAKISVPASEFAASLQGDDGQPYDKCQLLFDVND
ANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 52 proSap2 Mut3a Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDQNVTCQVTYSQDTADFCQKQGTYPGSSASQDLNTPFKIG
YGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
5 YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELRLSGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
IIFAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGD
NFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 53 proSap2 Mut3b Sequence

10 MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDQNVTCQVTYSQDTADFCQKQGTYPGSSASQDLNTPFKIG
IGYGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAY
SLYLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELRLSGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLA
DQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANIL
15 GDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 54 proSap2 Mut4a Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDQNVDCQVTQDNATADFCQKQGTYPGSSASQDLNTPFKIG
20 GYGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
LYLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELRLSGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
QIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILG
DNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

25 SEQ ID NO: 55 proSap2 Mut4c Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDQNVDCQVTQDNATADFCQKQGTYPGSSASQDLNTPFKIG
YGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELRLSGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
30 IIFAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGD
NFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 56 proSap2 Mut4d Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
35 QVTYAADITVGSNNQKLNIVDTGSSDLWVPDQNVDCQVTYSQDNATFCQKQGTYPGSSASQDLNTPFKIG

YGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
IIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNF SKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGD
NFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHH

5

SEQ ID NO: 57 proSap2 Mut4e Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQV TYSDQNATADFCQKGTYPGSSASQDLNTPFK
IGYGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAY
10 SLYLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLA
DQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNF SKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANIL
GDNFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHH

SEQ ID NO: 58 proSap2 Mut5a Sequence

15 MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQV TYSDQTADFCQKGTYPGSDQNATKQDLNTP
FKIGYGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKN
AYSLYLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQD
LADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNF SKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVND
20 NILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHH

SEQ ID NO: 59 proSap2 Mut5b Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQV TYSDQTADFCQKGTYPGSDQNATASQDLNT
25 PFKIGYGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAK
NAYSLYLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQ
DLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNF SKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVND
ANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHH

30 SEQ ID NO: 60 proSap2 Mut6a Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQV TYSDQTADFCQKGTYPGSSASQDLNTPFKIG
YGDQNATSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAY
SLYLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLA

DQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANIL
GDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 61 proSap2 Mut6b Sequence

5 MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVITYSDQTADFCKQKGTYPGSSASQDLNTPFKIG
YGDGDQATSSSQGTLYKDTVGGVSIKNQVLADVDSTSIDQILGVGYKTNEAGGSYDNVPVTLKKQGVIAK
NAYSLYLNSPDAATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQ
DLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVND
10 ANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 62 proSap2 Mut7 Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVITYSDQTADFCKQKGTYPGSSASQDLNTPFKIG
15 YDGSSSQGTLYKDTVGGVSIKNQVLADVDSTSIDQILGVGYKTNEAGGDQATKYDNVPVTLKKQGVIAK
NAYSLYLNSPDAATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQ
DLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVND
ANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

20 SEQ ID NO: 63 proSap2 Mut8a Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVITYSDQTADFCKQKGTYPGSSASQDLNTPFKIG
YDGSSSQGTLYKDTVGGVSIKNQVLADVDSTSIDQILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
YLNSPDQATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLAD
25 QIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILG
DNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 64 proSap2 Mut8b Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
30 QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVITYSDQTADFCKQKGTYPGSSASQDLNTPFKIG
YDGSSSQGTLYKDTVGGVSIKNQVLADVDSTSIDQILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
YLNSPDQATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQ
DLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVND
ANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

35

SEQ ID NO: 65 proSap2 Mut9 Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVITYSDQTADFCKQKGTYPSPGSSASQDLNTPFKIG
YDGGSSSQGTLYKDTVGFGGVSIGNQVLADV DSTSIDQILGVGYKTNEAGGSYDNPVTLKKQGVIAKNAYSL
5 YLNSPDAATGQIIFGGVDNAKYSDQNATKLIAPVTSRELRLISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQD
LADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDA
NILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 66 proSap2 Mut10 Sequence

10 MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVITYSDQTADFCKQKGTYPSPGSSASQDLNTPFKIG
YDGGSSSQGTLYKDTVGFGGVSIGNQVLADV DSTSIDQILGVGYKTNEAGGSYDNPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTSDQNATKELRLISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQD
LADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDA
15 NILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 67 proSap2 Mut11 Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVITYSDQTADFCKQKGTYPSPGSSASQDLNTPFKIG
20 YDGGSSSQGTLYKDTVGFGGVSIGNQVLADV DSTSIDQILGVGYKTNEAGGSYDNPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTSDRELRLISKDNASVEVSGKTINTDNVDVLVNSGTTITYLQQD
LADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDA
NILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

25 SEQ ID NO: 68 proSap2 Mut12 Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVITYSDQTADFCKQKGTYPSPGSSASQDLNTPFKIG
YDGGSSSQGTLYKDTVGFGGVSIGNQVLADV DSTSIDQILGVGYKTNEAGGSYDNPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTSDRELRLISLGSVEVSDQNATKTINTDNVDVLVNSGTTITYLQQD
30 LADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDA
NILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 69 proSap2 Mut13 Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFSVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
35 QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVITYSDQTADFCKQKGTYPSPGSSASQDLNTPFKIG

YGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDQ NATVDVLVNSGTTITYLQQDL
ADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNF SKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDAN
ILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHHH

5

SEQ ID NO: 70 proSap2 Mut14a Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNDCQV TYSDQTADFCKQKGTYPGSSASQDLNTPFKIG
YGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
10 YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
IIKAFDQ NATQDSNGNSFYEVDCNLSGDVVFNF SKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILG
DNFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHHH

SEQ ID NO: 71 proSap2 Mut14b Sequence

15 MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNDCQV TYSDQTADFCKQKGTYPGSSASQDLNTPFKIG
YGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
IIKAFNGKDQNASNGNSFYEVDCNLSGDVVFNF SKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILG
20 DNFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHHH

SEQ ID NO: 72 proSap2 Mut14c Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNDCQV TYSDQTADFCKQKGTYPGSSASQDLNTPFKIG
25 YGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
IIKAFNGKLTQ NATNSFYEVDCNLSGDVVFNF SKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGD
NFLRSAYIVYDLDDNEISLAQVKYTSASSISAL THHHHHHHHHHH

30 SEQ ID NO: 73 proSap2 Mut14d Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNDCQV TYSDQTADFCKQKGTYPGSSASQDLNTPFKIG
YGDGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTS DRELISLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ

I I K A F N G K L T Q D S D Q N A S F Y E V D C N L S G D V V F N F S K N A K I S V P A S E F A A S L Q G D D G Q P Y D K C Q L L F D V N D A N I L G
D N F L R S A Y I V Y D L D D N E I S L A Q V K Y T S A S S I S A L T H H H H H H H H H H

SEQ ID NO: 74 proSap2 Mut14e Sequence

5 M K I K T G A R I L A L S A L T T M M F S A S A L A S T P T T T K R S A G F V A L D F S V V K T P K A F P V T N G Q E G K T S K R Q A V P V T L H N E
Q V T Y A A D I T V G S N N Q K L N V I V D T G S S D L W V P D V N V D C Q V T Y S D Q T A D F C K Q K G T Y D P S G S S A S Q D L N T P F K I G
Y G D G S S S Q G T L Y K D T V G F G G V S I K N Q V L A D V D S T S I D Q G I L G V G Y K T N E A G G S Y D N V P V T L K K Q G V I A K N A Y S L
Y L N S P D A A T G Q I I F G G V D N A K Y S G S L I A L P V T S D R E L R I S L G S V E V S G K T I N T D N V D V L V N S G T T I T Y L Q Q D L A D Q
I I K A F N G K L T Q D S D Q N A S N G S F Y E V D C N L S G D V V F N F S K N A K I S V P A S E F A A S L Q G D D G Q P Y D K C Q L L F D V N D A N
10 I L G D N F L R S A Y I V Y D L D D N E I S L A Q V K Y T S A S S I S A L T H H H H H H H H H H

SEQ ID NO: 75 proSap2 Mut15 Sequence

M K I K T G A R I L A L S A L T T M M F S A S A L A S T P T T T K R S A G F V A L D F S V V K T P K A F P V T N G Q E G K T S K R Q A V P V T L H N E
Q V T Y A A D I T V G S N N Q K L N V I V D T G S S D L W V P D V N V D C Q V T Y S D Q T A D F C K Q K G T Y D P S G S S A S Q D L N T P F K I G
15 Y G D G S S S Q G T L Y K D T V G F G G V S I K N Q V L A D V D S T S I D Q G I L G V G Y K T N E A G G S Y D N V P V T L K K Q G V I A K N A Y S L
Y L N S P D A A T G Q I I F G G V D N A K Y S G S L I A L P V T S D R E L R I S L G S V E V S G K T I N T D N V D V L V N S G T T I T Y L Q Q D L A D Q
I I K A F N G K L T Q D S N G S F Y E V D C N L D Q N A S G D V V F N F S K N A K I S V P A S E F A A S L Q G D D G Q P Y D K C Q L L F D V N D A
N I L G D N F L R S A Y I V Y D L D D N E I S L A Q V K Y T S A S S I S A L T H H H H H H H H H H

20 SEQ ID NO: 76 proSap2 Mut16 Sequence

M K I K T G A R I L A L S A L T T M M F S A S A L A S T P T T T K R S A G F V A L D F S V V K T P K A F P V T N G Q E G K T S K R Q A V P V T L H N E
Q V T Y A A D I T V G S N N Q K L N V I V D T G S S D L W V P D V N V D C Q V T Y S D Q T A D F C K Q K G T Y D P S G S S A S Q D L N T P F K I G
Y G D G S S S Q G T L Y K D T V G F G G V S I K N Q V L A D V D S T S I D Q G I L G V G Y K T N E A G G S Y D N V P V T L K K Q G V I A K N A Y S L
Y L N S P D A A T G Q I I F G G V D N A K Y S G S L I A L P V T S D R E L R I S L G S V E V S G K T I N T D N V D V L V N S G T T I T Y L Q Q D L A D Q
25 I I K A F N G K L T Q D S N G S F Y E V D C N L S G D V V F N F S K D Q N A K I S V P A S E F A A S L Q G D D G Q P Y D K C Q L L F D V N D A
N I L G D N F L R S A Y I V Y D L D D N E I S L A Q V K Y T S A S S I S A L T H H H H H H H H H H

SEQ ID NO: 77 proSap2 Mut17 Sequence

M K I K T G A R I L A L S A L T T M M F S A S A L A S T P T T T K R S A G F V A L D F S V V K T P K A F P V T N G Q E G K T S K R Q A V P V T L H N E
30 Q V T Y A A D I T V G S N N Q K L N V I V D T G S S D L W V P D V N V D C Q V T Y S D Q T A D F C K Q K G T Y D P S G S S A S Q D L N T P F K I G
Y G D G S S S Q G T L Y K D T V G F G G V S I K N Q V L A D V D S T S I D Q G I L G V G Y K T N E A G G S Y D N V P V T L K K Q G V I A K N A Y S L
Y L N S P D A A T G Q I I F G G V D N A K Y S G S L I A L P V T S D R E L R I S L G S V E V S G K T I N T D N V D V L V N S G T T I T Y L Q Q D L A D Q
I I K A F N G K L T Q D S N G S F Y E V D C N L S G D V V F N F S K N A K I S V P A S E F A A S L Q G D D Q N A T Q P Y D K C Q L L F D V N D A N I
L G D N F L R S A Y I V Y D L D D N E I S L A Q V K Y T S A S S I S A L T H H H H H H H H H H

35

SEQ ID NO: 78 proSap2 Mut18 Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPGSGSSASQDLNTPFKIG
YDGGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNPVTLKKQGVIAKNAYSL
5 YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTSDRELRLSGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
IIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDQNATA
NILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 79 proSap2 Mut19 Sequence

10 MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPGSGSSASQDLNTPFKIG
YDGGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTSDRELRLSGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
IIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGD
15 NFLRSAYIVYDLQDNATEISLAQVKYTSASSISALTHHHHHHHHHH

SEQ ID NO: 80 proSap2 Mut20 Sequence

MKIKTGARILALSALTMMFSASALASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRQAVPVLHNE
QVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPGSGSSASQDLNTPFKIG
20 YDGGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNPVTLKKQGVIAKNAYSL
YLNSPDAATGQIIFGGVDNAKYSGSLIALPVTSDRELRLSGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQ
IIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGD
NFLRSAYIVYDLDDNEISLAQVKYTS DQNATISALTHHHHHHHHHH

25 SEQ ID NO: 81 proSap2 MutN3C-C Sequence

MIKFLSALILLVTTAAQASTPTTTKRSAGFVALDFS VVKTPKAFPVTNGQEGKTSKRLEVLFGPGGDQ NATKG
GGQAVPVLHNEQVTYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPGSGS
SASQDLNTPFKIGYDGGSSSQGTLYKDTVGFGGVSIKNQVLADV DSTSIDQGILGVGYKTNEAGGSYDNPVTL
KKQGVIAKNAYS LYLNSPDAATGQIIFGGVDNAKYSGSLIALPVTSDRELRLSGSVEVSGKTINTDNVDVLVNSG
30 TTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGDDGQPYDKC
QLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTGGGGGKDQ NATGGGGGHHHHHHHHH
H

SEQ ID NO: 82 proSap2 MutN3C-8a-C Sequence

MIKFLSALILLVTTAAQASTPTTTKRSAGFVALDFSVMKTPKAFPVTNGQEGKTSKRLEVLFGPGGDDQNATKG
GGQAVPVLHNEQVTYAADITVGSNNQKLNIVDTGSSDLWVPDNNVDCQVTYSDQTADFCKQKGTYPSPGS
SASQDLNTPFKIGYGDGSSSQGTLYKDTVGGVSIKNQVLADVDSTSIDQGILGVGYKTNEAGGSYDNVPVTL
5 KKQGVIAKNAYSILNSPDQATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNS
GTTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGGDGPYDK
CQLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTGGGGGKDQATGGGGGHHHHHHHH
HH

10 SEQ ID NO: 83 proSap2 MutN3C-3a-8a-C Sequence

MIKFLSALILLVTTAAQASTPTTTKRSAGFVALDFSVMKTPKAFPVTNGQEGKTSKRLEVLFGPGGDDQNATKG
GGQAVPVLHNEQVTYAADITVGSNNQKLNIVDTGSSDLWVPDQNVTCQVTYSDQTADFCKQKGTYPSPGS
SASQDLNTPFKIGYGDGSSSQGTLYKDTVGGVSIKNQVLADVDSTSIDQGILGVGYKTNEAGGSYDNVPVTL
KKQGVIAKNAYSILNSPDQATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNS
15 GTTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGGDGPYDK
CQLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTGGGGGKDQATGGGGGHHHHHHHH
HH

SEQ ID NO: 84 proSap2 MutN8a-fba-C Sequence

20 MIKFLSALILLVTTAAQASTPTTTKRSAGFVALDFSVMKTPKAFPVTNGQEGKTSKRQAVPVLHNEQVTYAADI
TVGSNNQKLNIVDTGSSDLWVPDNNVDCQVTYSDQTADFCKQKGTYPSPGSSASQDLNTPFKIGYGDGSSSQ
GTLYKDTVGGVSIKNQVLADVDSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYSILNSPDQN
ATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSGTTITYLQQDLADQIIKAFNG
KLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGGDGPYDKCQLLFDVNDANILGDNFLRSA
25 YIVYDLDDNEISLAQVKYTSASSISALTYGKDVKDLFDYAQEGGGGDDQNATGGGGGHHHHHHHHHH

SEQ ID NO: 85 proSap2 MutN3C-fba-C Sequence

MIKFLSALILLVTTAAQASTPTTTKRSAGFVALDFSVMKTPKAFPVTNGQEGKTSKRLEVLFGPGGDDQNATKG
GGQAVPVLHNEQVTYAADITVGSNNQKLNIVDTGSSDLWVPDNNVDCQVTYSDQTADFCKQKGTYPSPGS
30 SASQDLNTPFKIGYGDGSSSQGTLYKDTVGGVSIKNQVLADVDSTSIDQGILGVGYKTNEAGGSYDNVPVTL
KKQGVIAKNAYSILNSPDAATGQIIFGGVDNAKYSGLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSG
TTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNFSSKNAKISVPASEFAASLQGGDGPYDKC
QLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTYGKDVKDLFDYAQEGGGGDDQNATGG
GGGHHHHHHHHHH

SEQ ID NO: 86 proSap2 MutN3C-8a-fba-C Sequence

MIKFLSALILLVTTAAQASTPTTTKRSAGFVALDFSVMKTPKAFPVTNGQEGKTSKRLEVLFGPGGDQ⁵NATKG
GGQAVPVLHNEQV¹⁰TYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPSPGS
SASQDLNTPFKIGYGDGSSSQGTLYKDTVGFGGVSIKNQVLADVDSTSIDQGILGVGYKTNEAGGSYDNVPVTL
5 KKQGVIAKNAYS¹⁵LYLN²⁰SPDQ²⁵NATGQ³⁰IIFGGVDNAKYSGLIALPVTSDRELISLGSVEVSGKTINTDNVDVLVNS
GTTITYLQQLADQ³⁵IIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNF⁴⁰SKNAKISVPASEFAASLQGGDGPYDK
CQLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTYGKDVKDLFDYAQEGGGGGDQ⁴⁵NATG
GGGGHHHHHHHHHH

10 SEQ ID NO: 87 proSap2 MutN3C-3a-8a-fba-C Sequence

MIKFLSALILLVTTAAQASTPTTTKRSAGFVALDFSVMKTPKAFPVTNGQEGKTSKRLEVLFGPGGDQ¹⁵NATKG
GGQAVPVLHNEQV²⁰TYAADITVGSNNQKLNIVDTGSSDLWVPDQNVTCQVTYSDQTADFCKQKGTYPSPGS
SASQDLNTPFKIGYGDGSSSQGTLYKDTVGFGGVSIKNQVLADVDSTSIDQGILGVGYKTNEAGGSYDNVPVTL
KKQGVIAKNAYS²⁵LYLN³⁰SPDQ³⁵NATGQ⁴⁰IIFGGVDNAKYSGLIALPVTSDRELISLGSVEVSGKTINTDNVDVLVNS
15 GTTITYLQQLADQ⁴⁵IIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNF⁵⁰SKNAKISVPASEFAASLQGGDGPYDK
CQLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTYGKDVKDLFDYAQEGGGGGDQ⁵⁵NATG
GGGGHHHHHHHHHH

SEQ ID NO: 88 Wild type mature Sap2 sequence from *Candida albicans* (lacking the wild type

20 Leader sequence and lacking the wild type propeptide sequence; the amino acid residue (D) at
position 218 is double underlined; position 218 corresponds to position 274 of SEQ ID NO: 1)
(amino acid residues 57-398 of SEQ ID NO: 1)

QAVPVLHNEQV²⁵TYAADITVGSNNQKLNIVDTGSSDLWVPDVNVDCQVTYSDQTADFCKQKGTYPSPGSSAS
QDLNTPFKIGYGDGSSSQGTLYKDTVGFGGVSIKNQVLADVDSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQ
25 G³⁰VI³⁵AKNAYS⁴⁰LYLN⁴⁵SPDAATGQ⁵⁰IIFGGVDNAKYSGLIALPVTSDRELISLGSVEVSGKTINTDNVDVLVDSGTIT
YLQQLADQ⁵⁵IIKAFNGKLTQDSNGNSFYEVDCNLSGDVVFNF⁶⁰SKNAKISVPASEFAASLQGGDGPYDKCQLLF
DVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALT

30 SEQ ID NO: 89 Modified Sap2 sequence (histidine-tagged, intermediate form of Sap2-Fba fusion
protein lacking the native leader sequence, comprising the pro-peptide sequence (underlined);
comprising 4 glycosites (bolded), comprising an inactivating substitution (double underlined), and
further comprising an Fba sequence (dashed undelined)) (artificial sequence)

STPTTTKRSAGFVALDFSVMKTPKAFPVTNGQEGKTSKR **GGGDQ³⁵NAT** GGQAVPVLHNEQV⁴⁰TYAADITVGSN
NQLNIVDTGSSDLWVP**DQNV**TCQVTYSDQTADFCKQKGTYPSPGSSASQDLNTPFKIGYGDGSSSQGTLY
35 KDTVGFGGVSIKNQVLADVDSTSIDQGILGVGYKTNEAGGSYDNVPVTLKKQGVIAKNAYS⁴⁵LYLN**SPDQ⁵⁰NAT**G

QIIFGGVDNAKYSGSLIALPVTSDRELRLSLGSVEVSGKTINTDNVDVLVNSSGTTITYLQQDLADQIIKAFNGKLT
QDSNGNSFYEVDCNLSGDDVFNFSKNAKISVPASEFAASLQGDDGQPYDKCQLLFDVNDANILGDNFLRSAYIV
YDLDDNEISLAQVKYTSASSISALTYGKDVKDLFDYAQEGSGGG**DQ**NATGSGGGHHHHHHHHHH

- 5 SEQ ID NO: 90 Modified Sap2 sequence (histidine-tagged, mature form of Sap2-Fba fusion protein lacking the native leader sequence or the pro-peptide sequence, comprising 4 glycosites (bolded), comprising an inactivating substitution (double underlined), and further comprising an Fba sequence (dashed undelined) (artificial sequence)

GGG**DQ**NATGGGQAVPVTLHNEQVTYAADITVGSNNQKLNIVDTGSSDLWVP**DQ**NVTCQVTYSQDTADFC
10 KQKGTYPGSSASQDLNTPFKIGYGDGSSSQGTLYKDTVGFGGVSIKNQVLADVSTSIDQGILGVGYKTNEA
GGSYDNVPVTLKKQGVIAKNAYSLLNSP**DQ**NATGQIIFGGVDNAKYSGSLIALPVTSDRELRLSLGSVEVSGKT
INTDNVDVLVNSSGTTITYLQQDLADQIIKAFNGKLTQDSNGNSFYEVDCNLSGDDVFNFSKNAKISVPASEFAAS
LQGDDGQPYDKCQLLFDVNDANILGDNFLRSAYIVYDLDDNEISLAQVKYTSASSISALTYGKDVKDLFDYAQEG
SGGG**DQ**NATGSGGGHHHHHHHHHH

15

Consensus sequence (artificial sequence)
J-U-B-D/E-X-N-Z-S/T-J-U-B

Consensus sequence (artificial sequence)

- 20 K-D/E-X-N-Z-S/T-K

CLAIMS

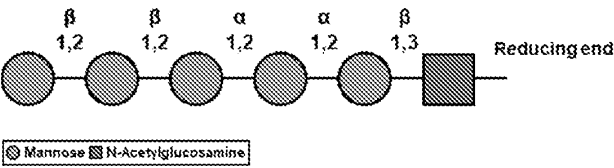
1. A modified Secreted Aspartyl Proteinase 2 (Sap2) protein comprising amino acid residues 19-398 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1, modified
5 in that the amino acid sequence comprises one or more consensus sequences comprising an amino acid sequence of D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.
2. The modified Sap2 protein of claim 1, wherein the protein further comprises a substitution at amino acid residue 274 of amino acid residues 19-398 of SEQ ID NO: 1 or at an
10 equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1, optionally wherein the protein comprises an Aspartic Acid (D) to Asparagine (N) substitution at amino acid residue 274 of amino acid residues 19-398 of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%,
15 95%, 96%, 97%, 98% or 99% identical to amino acid residues 19-398 of SEQ ID NO: 1.
3. The modified Sap2 protein of any of claim 1 and 2, wherein the modified Sap2 protein is from *Candida albicans*, and wherein the modified Sap2 protein further comprises at least one Fructose biphosphate aldolase-1 (Fba) peptide comprising an amino acid sequence of YGKDVKDLFDYAE (SEQ ID NO: 3) or an amino acid sequence at least 70%, 80%, 85%,
20 90%, or 92% identical to SEQ ID NO: 3.
4. The modified Sap2 protein of any of claims 1 to 3, wherein the one or more consensus sequences are selected from the group consisting of DQNAT (SEQ ID NO: 4) and DQNVN (SEQ ID NO: 5), and wherein the modified Sap2 protein comprises at least one additional consensus sequence comprising an amino acid sequence of J-U-B-D/E-X-N-Z-S/T-J-U-B,
25 wherein X and Z are independently any amino acid except proline and J, U and B independently comprise 1 to 5 naturally occurring amino acid residues, wherein the at least one consensus sequence has been added next to C-terminal amino acid residue of SEQ ID NO: 3 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 3.
- 30 5. A modified Sap2 protein comprising an amino acid sequence of SEQ ID NO: 9 or SEQ ID NO: 10.

6. A conjugate comprising the modified Sap2 protein of any of claims 1 to 5 and at least one saccharide antigen, optionally wherein the conjugate is a bioconjugate.
- 5 7. A modified Sap2 protein of *Candida albicans* comprising: (1) an amino acid sequence of SEQ ID NO: 9 or SEQ ID NO: 10; and (2) at least one saccharide chain from *Candida*, wherein the at least one saccharide chain is a β -1,2 mannan polymer consisting of at least five consecutive β -1,2 linked mannose molecules, and wherein the at least one saccharide chain is attached to at least one of four asparagine residues at positions 45, 94, 215, and 415 of SEQ ID NO: 9 or at least one of four asparagine residues at positions 6, 55, 176, and 376 of SEQ ID NO: 10.
- 10 8. A polynucleotide sequence encoding the modified Sap2 protein of any of claims 1 to 5.
9. A vector comprising the polynucleotide sequence of claim 8.
- 15 10. A host cell comprising: (1) one or more polynucleotide sequences that encode one or more heterologous glycosyltransferases; (2) a polynucleotide sequence that encodes a heterologous oligosaccharyl transferase; (3) a polynucleotide sequence that encodes a modified Sap2 protein according to any of claims 1 to 5; and, optionally, (4) a polynucleotide sequence that encodes a polymerase.
11. An immunogenic composition comprising the modified Sap2 protein of any of claims 1 to 5, the conjugate of claim 6, or the bioconjugate of claim 6.
- 20 12. A *Candida albicans* vaccine comprising: (1) the modified Sap2 protein of any of claims 1 to 5; (2) at least one *Candida albicans* saccharide linked to said modified Sap2 protein; and, optionally, (3) a pharmaceutically acceptable carrier or adjuvant.
- 25 13. A method for treatment or prevention of *Candida albicans* infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of the modified Sap2 protein of any of claims 1 to 5 and 7, the conjugate of claim 6, the bioconjugate of claim 6, the immunogenic composition of claim 11, or the vaccine of claim 12.
14. A method for immunizing a subject against *Candida albicans* infection, the method comprising administering to the subject an immunoprotective dose of the modified Sap2

protein of any of claims 1 to 5 and 7, the conjugate of claim 6, the bioconjugate of claim 6, the immunogenic composition of claim 11, or the vaccine of claim 12.

15. A method for inducing immune response to *Candida albicans* infection in a subject, the method comprising administering to the subject a therapeutically or prophylactically effective amount of modified Sap2 protein of any of claims 1 to 5 and 7, the conjugate of claim 6, the bioconjugate of claim 6, the immunogenic composition of claim 11, or the vaccine of claim 12.
16. The modified Sap2 protein of any of claims 1 to 5 and 7, the conjugate of any of claim 6, the bioconjugate of claim 6, the immunogenic composition of claim 11, or the vaccine of claim 12 for use in treatment or prevention of a disease caused by *Candida albicans* infection.

17. A saccharide that is a β -1,2 mannan polymer comprising the structure:



18. A saccharide comprising the structure: β -D-Man β -(1 $\rightarrow$ 2)- α -D-Man β -(1 $\rightarrow$ 2)- α -D-Man β -(1 $\rightarrow$ 2)- β -D-Man β -(1 $\rightarrow$ 3)-x-D-Glc β NAc.

19. A host cell comprising:
- i. a nucleotide sequence encoding one or more first heterologous glycosyltransferase(s) capable of synthesizing a β -1,2 mannan polymer;
 - ii. a nucleotide sequence encoding a second heterologous glycosyltransferase which is eukaryotic and is capable of covalently bonding a mannose molecule to a β -1,2 mannan polymer to extend a β -1,2 mannan polymer chain;
 - iii. a nucleotide sequence encoding a heterologous oligosaccharyl transferase; and
 - iv. optionally, a nucleotide sequence encoding a modified carrier protein comprising a glycosylation site comprising a consensus sequence D/E-X-N-Z-S/T, wherein X and Z are independently any amino acid except proline.

20. A method of producing a glycoconjugate comprising a modified carrier protein and a β -1,2 mannan, wherein said method comprises the steps of i) culturing the host cell of claim 19 under conditions suitable for the production of proteins, ii) harvesting the culture to produce a harvested culture, and iii) isolating the glycoconjugate from the culture.