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

(19) **World Intellectual Property Organization**
International Bureau

PCT

(43) **International Publication Date**
19 March 2009 (19.03.2009)

(10) **International Publication Number**
WO 2009/034041 A2

(51) **International Patent Classification:**
A61K 38/01 (2006.01) *C07K 14/435* (2006.01)

(21) **International Application Number:**
PCT/EP2008/061824

(22) **International Filing Date:**
5 September 2008 (05.09.2008)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
0717864.3 13 September 2007 (13.09.2007) GB

(71) **Applicant** *(for all designated States except US)*: **PEPTCELL LIMITED** [GB/GB]; 100 Fetter Lane, London Greater London EC4A 1BN (GB).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **STOLOFF, Gregory Alan** [AU/GB]; 32 Royal Avenue, London Greater London SW3 4QE (GB). **CAPARROS-WANDERLEY, Wilson Romero** [ES/GB]; 30 Great Meadow Way, Aylesbury Buckinghamshire HP19 7GE (GB).

(74) **Agent: PAGE WHITE & FARRER**; Bedford House, John Street, London Greater London WC1N 2BF (GB).

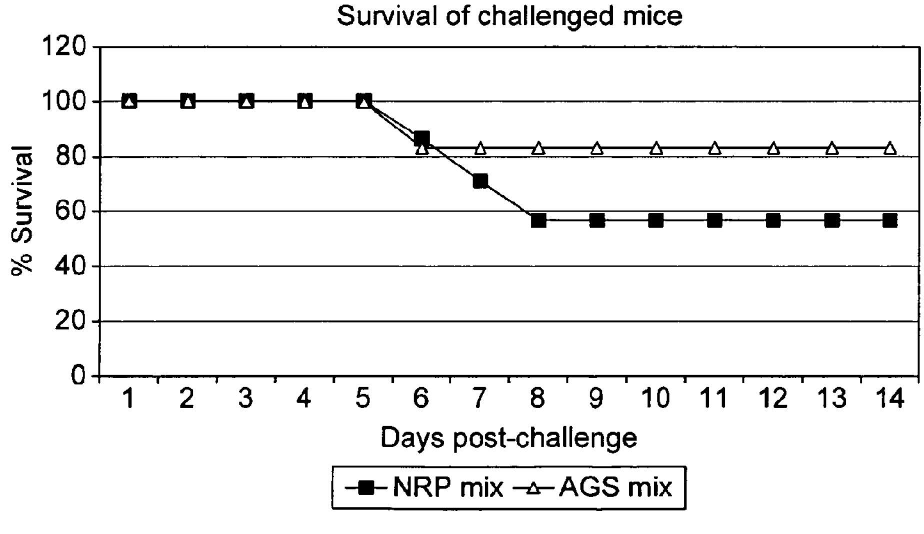
(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

(54) Title: PEPTIDE SEQUENCES AND COMPOSITIONS

FIG. 8



(57) **Abstract:** Provided is a polypeptide composition comprising one or more polypeptides, which polypeptides are immunogenic in a vertebrate such that they cause the vertebrate to produce immune system cells capable of recognising at least one epitope from an arthropod saliva protein fraction, wherein the arthropod saliva protein fraction has a mass of 4OkDA or less, and wherein the polypeptides are selected independently from: the polypeptide sequences of SEQ ID 1-44 or sub-sequences from these sequences, the subsequences having 7 amino acids or more; or from polypeptide sequences having 85% homology or more with one or more of the above sequences and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

WO 2009/034041 A2

PEPTIDE SEQUENCES AND COMPOSITIONS

The invention concerns peptide sequences, compositions comprising the peptide sequences, in particular vaccines against arthropod borne diseases comprising the sequences and the compositions, and uses of the sequences. The present invention is particularly concerned with vaccines that are protective against any one or more mosquito borne diseases, including one or more strains of malaria.

The defence against disease is critical for the survival of all animals, and the defence mechanism employed for this purpose is the animal immune system. Understanding the immune system is therefore a key to understanding the development of new and more sophisticated treatments for humans and animals alike.

The mechanism of operation of the immune system has been under investigation for many years. The system is composed of a number of cell types and a variety of molecules, making it extremely complex. Even after many years of study, the full extent of the immune system components, and their interaction with each other, is imperfectly understood.

Many years ago it was recognised that a person who recovers from a particular disease may acquire some protection in future against that disease, but not against a disease which that person has not yet contracted. This fundamental aspect of the immune system was interpreted at that time by considering that the immune system acquired a kind of 'memory' against certain pathogens once exposure to such pathogens had taken place, that memory being specific to a certain disease.

Gradually, it became known that exposure to less harmful variants of a pathogen could induce protection against more harmful variants (e.g. exposure to cowpox to protect against smallpox, or exposure to an inactivated anthrax to protect against live anthrax). Thus, the idea of vaccination against a disease arose.

It is now known that the immune system has at least two divisions: innate immunity and adaptive immunity. The innate system is fully functional before a pathogen enters the system, whilst the adaptive system is switched on after the pathogen enters the system. It then develops an attack specific to the pathogen. The innate system comprises a number of components, including phagocytes such as macrophages, which (as the name suggests) 'eat' or engulf foreign bodies such as pathogens.

Typically, but not exclusively, the present invention is concerned with the adaptive immune system, and unless specifically indicated otherwise, 'immune system' in the present context refers to the adaptive immune system.

In order to understand more fully how the immune system functions, the role of its individual components must be carefully considered. In respect of the adaptive immune system, it is well known that immunity against pathogens is provided by the action of lymphocytes, which constitute the most common cell type in the immune system. There are two types of lymphocyte: the B lymphocyte and the T lymphocyte. These are generally termed B cells and T cells respectively.

B cells have the ability to develop into plasma cells, which manufacture antibodies. Antibodies are very important components of the animal immune system. They are produced in response to some signature portion of the invading pathogen (an antigen of the pathogen – antigens here being defined as any foreign substance recognised by the immune system) and are usually specific to that pathogen. However, if two pathogens are very similar, or at least contain the same antigen, then antibodies produced against one can nevertheless be effective against the other (they may 'cross-react'). This explains why inoculation with cowpox may protect against smallpox. It is important to realise that the antibodies 'recognise' only a small portion of the antigenic molecule of the pathogen rather than the pathogen as a whole. These portions are termed epitopes.

T cells do not possess or produce antibodies. Instead, they recognise fragments (i.e. epitopes) of the foreign antigen complexed with major histocompatibility complex (MHC) (or in the case of humans, human leucocyte antigen (HLA)) via a specialised receptor known as TCR (T cell receptor). T cells are themselves divisible into subsets which can have either a regulatory function or an effector function. The effector cells are involved with 'effecting' the removal of foreign substances. For example, cytotoxic T cells (CTL) are effector cells that are able to kill infected cells, as well as other unwanted species such as tumour cells. Regulatory T cells, on the other hand, play a role in helping effector T and B cells to become more effective. Due to this function, these regulatory T cells are often termed 'helper' T cells. Other regulatory T cells, termed 'suppressor' T cells, are thought to inhibit immune responses, but these are less well understood. Regulatory T cells may also interact with components of the innate immune system to boost their activity.

In a normal healthy individual, the lymphocytes in the immune system remain in an inactive 'resting' state until an immune response is triggered. When an immune response is required, the lymphocytes become activated, proliferate and begin to carry out their designated functions. For example, any resting T cell displaying on its surface a TCR that recognises an epitope of the invading pathogen complexed with a MHC molecule is activated, proliferates (this being termed clonal expansion) and the resulting offspring start to actively carry out their predetermined effector functions required to combat the invading organisms.

When the immune response is completed, (i.e. the pathogens and/or infected cells have been eliminated) the lymphocytes revert to a resting state once again. This resting state is not, however, equivalent to the initial inactive resting state. Activated, but resting lymphocytes, can be rapidly recruited and induced to proliferate in response to an infection by the same, or closely related, pathogen at a later time.

This ability of activated resting lymphocytes, to deliver a faster and more powerful response following a second encounter with an invading pathogen, effectively provides the immune system with 'memory'. The exploitation of the immune system's memory is the basis for all long-term immunoprophylactic drugs (e.g. vaccines) and remains the goal of much long-term immunotherapeutic drug development.

In order for cells to perform their functions within the complex systems of an animal, the cells need to have 'receptors' on their surfaces. These receptors are capable of 'recognising' specific substances that control various essential processes such as activation, proliferation and adherence to other cells or substrates. For example, in the case of the immune system, the receptors on T and B cells allow them not only to recognise antigen but also to interact with each other and thus regulate their activities. Without these receptors, the cells would lack an essential means of communication and would be unable to act effectively in the concerted way that is essential for the immune system of a multicellular organism.

In order to be able to specifically recognise and deal with the wide range of pathogens present in the environment, the immune system has developed two types of highly variable antigen receptor on lymphocytes: antibodies in B cells and T cell receptors, or TCRs, in T cells.

There are a great many different possible antigen receptors present in the body, to enable the immune system to recognise a wide variety of invading pathogens. In fact there are approximately 10^{12} different B cells and T cell receptors in an individual. Each individual B cell has only one type of receptor, and so to deal with a particular pathogen, a B cell having the 'best fitting' receptor for an antigen of that pathogen must be selected. This process is termed 'clonal selection'. In theory, only a single clone may respond (a monoclonal response) or several (an oligoclonal response) or many (a polyclonal response) depending on the number of antigens/epitopes exhibited by the pathogen, and the specificity of the various selected B cells to these antigen/epitopes.

There is a major difference between the types of antigen that can be recognised by B cells and T cells. As far as it is known, only the receptors on the surface of B lymphocytes (i.e. antibodies) are capable of directly recognising antigens such as proteins on viruses and bacteria, or foreign molecules dissolved in body fluid. Antibodies can also be produced in a soluble form by the B cells when they are activated and develop into plasma cells. The antibodies are also termed immunoglobulins (abbreviated to Ig). T cell receptors, on the other hand, recognise only short peptides, also known as T cell epitopes, on the surface of cells of the body. These T-cell epitopes are produced by degradation of larger proteins that are either self (i.e. naturally occurring body proteins) or non-self (i.e. derived from foreign organisms infecting the body). Only those derived from foreign proteins, i.e. antigens, are normally capable of inducing an immune response in the body. Once produced, these epitopes are bound to a special type of molecule, the MHC (major histocompatibility complex) and the resulting complex is then presented on the cell surface for binding the T cell receptor.

It should be clear that due to the destructive nature of the immune response, the response has to act only against foreign pathogens, not against the body's own cells or proteins. Thus, the immune system needs to distinguish between 'self' and 'non-self'. It has been proposed that although clones of lymphocytes reacting against self are produced, they are deleted before any reaction can occur. This process is termed 'clonal deletion'. It has also been proposed that any self-reacting lymphocytes could be retained but only in a 'switched-off' state. This mechanism is termed 'clonal anergy'. Whatever the process considered, it remains unclear what is the exact underlying mechanism allowing lymphoid tissues, such as the thymus, to identify individual T cell clones reacting against self from the pool of T lymphocytes reacting only against non-self.

It has been known for many years that the major histocompatibility complex (MHC) plays a key role in the immune system of animals. The MHC molecules enable T cells to recognise antigens, as has already been discussed above. There are three general types of MHC molecule, class I, class II and class III. Class I and class II MHC molecules are glycoproteins that are present on the surface of the cell, whilst class III are usually soluble molecules present inside the cell. There are a large number of different types of MHC molecule. For example in humans (where MHC is termed HLA, human leukocyte antigen) there are several hundreds of different alleles of the genes coding for MHC molecules, meaning that in the human population there are many different types of HLA. The MHC of different species is typically named according to different conventions, thus MHC for mouse is termed H-2, for rat RT1 and for rabbit RLA. The different gene regions coding for different MHC molecules in an individual are usually individually named, such as HLA-A, HLA-C etc. in humans.

The MHC molecule is a critical immune system molecule, since it is this molecule that presents the epitopes of the antigens to the immune system. For example, if a T cell is to respond to a particular pathogen, the pathogen must have a least one antigen (such as a protein) that has at least one epitope (such as a peptide portion of the protein) that can bind to an MHC molecule on the surface of a cell and thus interact with a T cell which binds to the

MHC-peptide complex. Thus, the immune response is dependent on the ability of the MHC to bind to an epitope. If there is no epitope that the MHC will bind to, or if there is no T cell which will bind to the MHC-peptide complex, then no immune response will occur.

In respect of 'self' proteins, however, one of several epitopes may be able to bind to the MHC molecule and hence potentially induce an immune response. On these occasions a specific "signal" must be provided for the self-reacting lymphocyte clones to be deleted or "switched off".

Despite the growing knowledge of the workings of the vertebrate immune system, vaccines against many diseases remain elusive. Certain pathogens are subject to rapid mutation (e.g. HIV and influenza), such that epitopes that may be useful vaccine targets for one strain, are not useful after mutation has occurred because the epitope has changed in the new strain. Other pathogens such as plasmodium (the pathogen responsible for malaria) have been extensively investigated, but targets on the pathogen that can be useful in vaccine development have proven difficult to identify, or have simply failed to deliver an effective vaccine *in vivo*.

Pathogens for which there is a particular need for progress are those that are borne by arthropods and cause so-called "arthropod borne diseases". Such diseases are a major cause of death worldwide, and include *inter alia* malaria and dengue some of the greatest killers in the population today, particularly in poorer regions of the world. Examples include, but are not limited to, those listed in Table 1 below.

Table 1 - Arthropod-borne diseases, with their pathogens and vectors

Disease	**Pathogen (parasite/virus)**	**Arthropod Vector**
Malaria	*Plasmodium sp.*	*Anopheles sp.*
Filariasis	*Wuchereria sp.* *Brugia sp.* *Loa sp.* *Mansonella sp.*	*Anopheles sp.* *Aedes sp.* *Culex sp.* *Mansonia sp.* *Chrysops sp.*
River blindness	*Onchocerca sp.*	*Simulium sp.*
Chagas Disease Sleeping Sickness	*Trypanosoma sp.*	*Triatoma sp.* *Rhodnius sp.*

		Panstrongylus sp. *Glossina sp.* *Tabanus sp.*
Leishmaniasis	*Leishmania sp.*	*Phlebotomus sp.* *Lutzomyia sp.*
Tularaemia	*Franciscella tulariensis*	*Tabanus sp.*
Relapsing fever Lyme disease	*Borrelia sp.*	*Pediculus sp* *Ornithodoros sp.* *Ixodes sp.*
Typhus	*Rickettsia sp.*	*Pediculus sp.* *Pulex sp.*
Plague	*Yersinia sp.*	*Pulex sp.*
Dengue fever	*DEN-1 to -4 flavivirus*	*Aedes sp.*
Yellow fever	*Yellow fever flavivirus*	*Aedes sp.* *Hemagogous spegazzani.*
Rift valley fever	*Rift valley virus*	*Aedes sp.*
Encephalitis Including St. Louis Encephalitis West Nile fever Eastern Equine Encephalitis La Crosse Encephalitis Eastern Equine Encephalitis	*Flavi-, Alpha- and Bunyaviruses*	*Culex sp.* *Ochlerotatus sp.* *Aedes sp.* *Coquillettidia sp.* *Anopheles sp.*

Previously, attempts to provide arthropod borne disease vaccines have involved identifying existing pathogen strains and then producing a vaccine specific to that pathogen. Generally, the vaccines have been based upon a B cell (antibody) response (or occasionally a T cell response) the antibody being reactive with the surface antigens of the specific pathogen strain against which it has been developed. Typically, the surface proteins comprising the antigens are variable from one disease strain to the next, and completely different for different pathogens. The consequence of this is that conventional vaccines generally protect only against one specific pathogen or pathogen strain (if they are effective at all) and will not protect against other strains or a new strain that results from a mutation. Thus, a separate vaccine is required for protection against each disease and often against different and/or new strains of the same disease.

It has been known for some time that immunisation against arthropod saliva, or against antigens expressed in the gut of arthropods can protect an individual from infection. Articles discussing this include those summarised below.

R.G. Titus *et al.* "The immunomodulatory factors of arthropod saliva and the potential for these factors to serve as vaccine targets to prevent pathogen transmission.", *Parasite Immunology*, 2006, 28, 131-141. The article reviews known immunomodulators in arthropod salivary glands. Immunomodulatory factors can enhance pathogen transmission by arthropods. Vector saliva contains a large number of substances whose activities include the ability to inhibit haemostasis, vasoconstriction and the development of inflammation and an immune response. In mosquitoes there are several T-cell inhibitors. If arthropod saliva enhances infection with the pathogens arthropods transmit it should be possible to control pathogen transmission by vaccinating the host against the molecule(s) in saliva that potentiate the infection, thereby blocking the enhancing effects of saliva and thus preventing the pathogen from establishing infection in the host. The gene maxadilan or MAX that encodes the potent vasodilator in sandfly saliva was cloned and the activities of the protein product of the gene were examined. The effects of sandfly MAX manifest through primary effects on phagocytic cells and this leads to downstream immunomodulatory/inhibitory effects on T-cell responses. Mice were vaccinated against MAX by injecting MAX in complete Freund's adjuvant followed by MAX in incomplete Freund's adjuvant and then boosted with soluble MAX until circulating anti-MAX antibody titre of between 1/10000 and 1/20000. The vaccinated mice were markedly protected from infection.

G.A.T. Targett, "Parasites, arthropod vectors, and immune responses.", *Parasite Immunology*, 2006, 28, 117-119. This article shows that blood feeding by mosquitoes induces immune responses to insect's saliva. These include both hypersensitivity reactions and anti-mosquito antibodies. Measurement of these antibody responses can be of value epidemiologically for monitoring vector populations and the application of such responses to assessment of the effectiveness of intervention strategies. The production of vaccines that would impair the feeding, development and/or survival of arthropod vectors or ectoparasites is an attractive if ambitious idea. One important objective is control of the ectoparasitic arthropods themselves, while a second is to use the anti-arthropod vaccine as a way to harm the parasites that the arthropod transmits either within the arthropod itself or when transmission to the vertebrate host occurs.

M.M.A. Whitten *et al.*, "Mosquito midguts and malaria: cell biology, compartmentalization and immunology.", *Parasite Immunology*, 2006, 28, 121-130. The mosquito midgut represents one of the most challenging environments for the survival and development of Plasmodium. During their attempts to cross the midgut epithelium en route to the salivary glands, motile ookinetes are swiftly detected and labelled by mosquito recognition factors and targeted for destruction by a variety of immune responses that recruit killing factors both from the midgut and from other tissues in the surrounding body cavity. The exact interplay between these factors and the parasite is highly species and strain specific as are the timing and the route of the parasite invasion. The midgut forms a physical barrier separating and protecting the haemocoel tissues from digestive enzymes and infectious agents. It is composed of a single layer of polarised epithelial cells, with either pole displaying different morphological adaptations for increasing the surface area involved in the exchange of molecules. The distinct microvillous apical surface is exposed to the lumen, and its primary role is the secretion of digestive enzymes and absorption of nutrients. Structural modifications triggered by blood feeding include the formation of a thick non-cellular, chitinous peritrophic matrix (PM) that is secreted by the midgut epithelium and that is polymerized by the ingested blood meal. The PM surrounds the blood bolus and forms a barrier for parasites and bacteria attempting to penetrate the midgut epithelium.

P.F. Billingsley *et al.* "Immune interactions between mosquitoes and their hosts.", *Parasite Immunology*, 2006, 28, 143-153. Antibodies targeting the mosquito midgut are also important in the development of mosquito vaccines. The secretion of saliva by a mosquito during feeding is important for the successful location of host blood vessels and manipulation of host haemostatic and immune responses. Immunoblotting techniques have been used to characterise the recognition by host anti-mosquito antibodies of salivary antigens. When a mosquito feeds, it ingests host immune factors, both soluble and cellular, that remain active in the midgut. In contrast to salivary antigens, the host will not normally have been exposed to antigens from the internal organs of the mosquito, which leads to the idea of these 'concealed' antigens as vaccine targets. The best source of concealed antigens is the midgut, because after feeding it contains the blood meal with its component immune effector molecules and cells. Midguts and whole body preparations of *Ae. Aegypti* induced high antibody titres in mice and the increased mortality of mosquitoes that fed on them was correlated with antibody that bound to midgut microvilli. IgG responses were induced in mice injected with a midgut cDNA library of *An. Gambiae*, and high antibody titres achieved after boosting with midgut proteins. Mosquitoes feeding on these mice showed reproducible reductions in survival and fecundity, but interestingly cellular rather than humoral responses seemed to be responsible for the mosquitocidal effects. These studies have encountered problems. Even within experiments, variability can be high and reproducibility of effect difficult to achieve. Immunization with a complex mixture of either midgut protein extracts or a midgut cDNA library has meant that protective target antigens have not been identified.

However, despite this knowledge, no effective vaccine against arthropod borne diseases has yet been developed that makes use of this mechanism of action. A further significant problem with existing vaccines against arthropod borne pathogens, whether relying on a B-cell or T-cell response, is that each protects only against a single pathogen, or at most a single existing pathogen strain, and does not provide protection against possible future strains or against multiple pathogens. There is a desperate need for a vaccine to protect against a plurality of arthropod borne diseases, including large scale killers such as Malaria and Dengue.

The inventors have been now able to identify specific immunogenic peptide sequences present in arthropod saliva proteins that may provide protection against all arthropod borne pathogens, and have developed vaccines to the diseases they cause, using the identified sequences. Thus, the inventors have developed peptides useful in vaccines eliciting an immune response, and in particular a rapid secondary immune response against arthropod borne diseases.

Accordingly, it is an aim of the present invention to solve the problems associated with the known prior art as set out above. It is a further aim of the present invention to provide a polypeptide composition that is capable of eliciting a immune response (e.g. a cellular response including a T-cell response and/or a B cell response) in vertebrates against a plurality of arthropod borne diseases, i.e. diseases caused by a plurality of pathogens or pathogen strains. It is a further aim of the present invention to provide an arthropod borne disease vaccine using the polypeptide compositions of the invention.

Accordingly, the present invention provides a polypeptide composition comprising one or more polypeptides, which polypeptides are immunogenic in a vertebrate, such that they cause the vertebrate to produce immune system cells (or promote the production of such cells) capable of recognising at least one epitope from an arthropod saliva protein fraction, wherein the arthropod saliva protein fraction has a mass of 40kDA or less, and wherein the polypeptides are selected independently from:

(a) the polypeptide sequences of SEQ ID 1-44 or sub-sequences from these sequences, the sub-sequences having 7 amino acids or more:

SEQ ID 1	HLTLFTVAVLLLAAAALLLLLPPAYSTTLTPP
SEQ ID 2	PLSYCHLFLTHTLARALSFSRSDCL
SEQ ID 3	KNVFFALLLVVLVCCLVSVQGNEI
SEQ ID 4	KLLVLLICLFFYHTHCTTAYLWLAMGV
SEQ ID 5	FLKGSFPRFQMCVMLIGFFSSAKCL
SEQ ID 6	NDYQALLGLCCPWIDLAAADLPMRRHAKA
SEQ ID 7	FYSVGKLVKVLLVMAVCCLLLCTAPTGADPL
SEQ ID 8	MKFAFAFVLIALFAVFAVSQALPQPEQAAA
SEQ ID 9	DGASAITKIVLELTPEQAAAV
SEQ ID 10	TLFIFLVCCQIPLFGIMSSDSADPFYWIRVILA
SEQ ID 11	GRVMCLLRLMSTLLVVLSIVGK
SEQ ID 12	LYSGYRLLVLLVMTVCCLLLFIAPTGADPLPGQTQRTL
SEQ ID 13	MYCVIKGKTGGYCNSEGLCTCRAEDLHFLLKPIINKD
SEQ ID 14	NAEDPRTELIGCGSVLFHLAANRLSLQLEEFAVCKR
SEQ ID 15	ALIGLLLCSVQSVTANDPVDALGACSGNLFGLLMTRL
SEQ ID 16	SKLFVLAFLCLALVVVVQSAPQYARGDVPT
SEQ ID 17	SMLVAFATLSVALVVVVAIPANFNYGGGGGYFINGTGQ
SEQ ID 18	IYEKLPAYLSEVSARVNVLQVSLQHDLPNLQ
SEQ ID 19	EMKLAKVALVTISLWFMAWTPYLVINFTGI
SEQ ID 20	LLPAKVIPDKTAAYVAYGGQETLVEHVEVLV
SEQ ID 21	FYTCFLGTSSLAGFKNAVDYDELLKAG

SEQ ID 22	VLEVLGFVEDNGELVFQELLGVLKMVDPDGD
SEQ ID 23	KLTPTVVVVLLCLTFVADALTIQELRAQIAQQRIQQRYGVTVATT
SEQ ID 24	SLSDYGLIELKEHCLECCQKDTEADSKLKVYPAAVLEV
SEQ ID 25	TYICFILHGVSEIIPQQQKKTMKFLLLVASVLCLVLI
SEQ ID 26	RYFVVIALICPLIIVETLAV
SEQ ID 27	LLLYLDAADLRRALHQYQLLAAQGDRHLPQQIVKFV
SEQ ID 28	VLLTPALQAYIMDEHNLNRSNIALGRIRPYPSAVKMP
SEQ ID 29	VLKGETHKALKLKDGGHYLVEFKSIYM
SEQ ID 30	VLHSMLVNASLAEMVKESYQTHGADGRMVVRMLKFVRLLP
SEQ ID 31	RVRALRALLETLLQHQGEQNNDVYLIRLAHET
SEQ ID 32	ELQQALSSLNAGSGSCAEVFNAYLPVHNKYIGVSRKI
SEQ ID 33	KFYRLISTLLVVVVIAPRHQCSPFFFQYNRPYL
SEQ ID 34	NYVPDVSALEQDIIEVDPETKEMLKHLDFNNIVVQL
SEQ ID 35	QYSMECLEAAEPKYLDGLKTLADETAQC
SEQ ID 36	EYAQVTKMLGNGRLEAMCFDGVKRLCHIRGKL
SEQ ID 37	KLFLTLLSTLSVAMVFALPAHHHSRG
SEQ ID 38	ELEEARLVAEELEERQQELDYLKRYLVGRLQAV
SEQ ID 39	SYFLTVCLLALVQSETVQD
SEQ ID 40	AMTNANLVGLTISLAYAIFFLLYTPPTGRSS
SEQ ID 41	SFAWLLYGIILRSNFLVVQNLMALALSAVQLSLFII
SEQ ID 42	AFPFISGFLSCFMWLKYGVLTEESTLILVNFIGSAL
SEQ ID 43	GLLCCCLAVLFFASPLTMLAHVIR
SEQ ID 44	LLLAMVLLPLLLLESVVPYAAAEKVW

(b) the sequences defined by the following amino acid residues of an arthropod saliva protein, or sub-sequences from these sequences, the sub-sequences having 7 amino acids or more:

residues 2-33 of >gi|18389913|gb|AAL68793.1|AF457563_1 hypothetical protein 16 [Anopheles gambiae]

residues 2-26 of >gi|18389909|gb|AAL68791.1|AF457561_1 hypothetical protein 14 [Anopheles gambiae]

residues 2-25 of >gi|18389907|gb|AAL68790.1|AF457560_1 hypothetical protein 13 [Anopheles gambiae]

residues 10-36 of >gi|18389903|gb|AAL68788.1|AF457558_1 hypothetical protein 11 [Anopheles gambiae]

residues 2-26 of >gi|62546227|gb|AAX86005.1| hyp3.5 precursor [Anopheles gambiae]

residues 14-42 of gi|18389899|gb|AAL68786.1|AF457556_1 salivary gland 7-like protein [Anopheles gambiae]

residues 3-33 of >gi|18389911|gb|AAL68792.1|AF457562_1 hypothetical protein 15 [Anopheles gambiae]

residues 1-30 of >gi|62546225|gb|AAX86004.1| hyp6.3 precursor [Anopheles gambiae]

residues 34-54 of >gi|62546225|gb|AAX86004.1| hyp6.3 precursor [Anopheles gambiae]

residues 38-70 of >gi|17026153|emb|CAD12038.1| Sec61 protein [Anopheles gambiae]

residues 2-23 of >gi|62546223|gb|AAX86003.1| hyp6.2 precursor [Anopheles gambiae]

residues 17-54 of >gi|18389915|gb|AAL68794.1|AF457564_1 hypothetical protein 17 [Anopheles gambiae]

residues 57-93 of >gi|87080391|gb|ABD18596.1| defensin [Anopheles gambiae]

residues 22-57 of >gi|18389901|gb|AAL68787.1|AF457557_1 hypothetical protein 10 [Anopheles gambiae]

residues 7-43 of >gi|18389905|gb|AAL68789.1|AF457559_1 hypothetical protein 12 [Anopheles gambiae]

residues 3-32 of >gi|4127344|emb|CAA76832.1| cE5 protein [Anopheles gambiae]

residues 3-40 of >gi|4210617|emb|CAA10259.1| SG2 protein [Anopheles gambiae]

residues 91-121 of >gi|4127309|emb|CAA76820.1| hypothetical protein [Anopheles gambiae]

residues 65-94 of >gi|4375824|emb|CAA76825.1| opsin [Anopheles gambiae]

residues 41-71 of >gi|62546233|gb|AAX86008.1| unknown [Anopheles gambiae]

residues 117-143 of >gi|3378531|emb|CAA03872.1| D7r2 protein [Anopheles gambiae]

residues 63-93 of >gi|3378529|emb|CAA03871.1| D7r3 protein [Anopheles gambiae]

residues 23-67 of >gi|18389893|gb|AAL68783.1|AF457553_1 mucin-like protein [Anopheles gambiae]

residues 43-80 of >gi|18389881|gb|AAL68777.1|AF457547_1 selenoprotein [Anopheles gambiae]

residues 6-42 of >gi|18389879|gb|AAL68776.1|AF457546_1 30 kDa protein [Anopheles gambiae]

residues 4-23 of >gi|18378603|gb|AAL68639.1|AF458073_1 D7-related 5 protein [Anopheles gambiae]

residues 20-55 of >gi|18389897|gb|AAL68785.1|AF457555_1 salivary gland 1-like 4 protein [Anopheles gambiae]

residues 59-95 of >gi|18389883|gb|AAL68778.1|AF457548_1 antigen 5-related 1 protein [Anopheles gambiae]

residues 158-184 of >gi|83016748|dbj|BAE53441.1| DsRed [synthetic construct]

residues 37-76 of >gi|18389895|gb|AAL68784.1|AF457554_1 salivary gland 1-like 3 protein [Anopheles gambiae]

residues 191-222 of >gi|18389895|gb|AAL68784.1|AF457554_1 salivary gland 1-like 3 protein [Anopheles gambiae]

residues 113-149 of >gi|18389891|gb|AAL68782.1|AF457552_1 D7 protein long form [Anopheles gambiae]

residues 1-37 of >emb|CAC35527.1| gSG9 protein [Anopheles gambiae]

residues 81-120 of >sp|Q9U9L1|RS17_ANOGA 40S ribosomal protein S17

residues 111 to 142 of >emb|CAC35523.1| gSG7 protein [Anopheles gambiae]

residues 32-67 of >gb|AAD47075.1|AF164151_1 translation initiation factor 4C (1A) [Anopheles gambiae]

residues 1-29 of >emb|CAC35519.1| gSG2-like protein [Anopheles gambiae]

residues 106-142 of >emb|CAC35451.1| hypothetical protein [Anopheles gambiae]

residues 6-28 of > emb|CAC35524.1| D7r4 protein [Anopheles gambiae]

residues 70-104 of >ref|XP_001230998.1| ENSANGP00000014906 [Anopheles gambiae str. PEST]

residues 174-213 of >ref|XP_316361.2| ENSANGP00000012984 [Anopheles gambiae str. PEST]

residues 41-80 of >ref|XP_314140.3| ENSANGP00000015780 [Anopheles gambiae str. PEST]

residues 126-153 of >ref|XP_314140.3| ENSANGP00000015780 [Anopheles gambiae str. PEST]

residues 5-34 of >emb|CAC35522.1| gSG6 protein [Anopheles gambiae]

(c) and from polypeptide sequences having 85% homology or more with one or more of the sequences in (a) or (b) and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

CDS is short for "CoDing Sequence", i.e. a region of nucleotides that corresponds to the sequence of amino acids in a predicted protein. The CDS includes start and stop codons, therefore coding sequences begin with an "ATG" and end with a stop codon. Unexpressed sequences, including the 5'-UTR, the 3'-UTR, introns, or bases not expressed due to frameshifting, are not included within a CDS. Note that the CDS does not correspond to the actual mRNA sequence. As a result, a CDS translation is the protein that would result if all the codons between the start and stop codons were translated.

PDB stands for Protein Data Bank. This database (http://www.rcsb.org/pdb/home/home.do) is maintained by the Research Collaboratory for Structural Bioinformatics (RCSB), a non-profit consortium dedicated to improving the understanding of the function of biological systems through the study of the 3-D structure of biological macromolecules.

The Protein Information Resource (PIR) (http://pir.georgetown.edu/), is an integrated public bioinformatics resource established in 1984 by the National Biomedical Research Foundation (NBRF) as a resource to assist researchers in the identification and interpretation of protein sequence information (Wu CH, Yeh LS, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu Z, Kourtesis P, Ledley RS, Suzek BE, Garrett L, Vinayaka CR, Zhang J, Barker WC (2003). "The Protein Information Resource". *Nucleic Acids Res 31(1):345-7.*)

PRF is an online database maintained by the Protein Research Foundation (PRF) (http://www.prf.or.jp/en/index.shtml). The database contains information related to amino acids, peptides and proteins collected from scientific journals, peptide and Protein sequence data, data on synthetic compounds and molecular aspects of proteins.

GenBank is the NIH genetic sequence database (http://www.ncbi.nlm.nih.gov/), an annotated collection of all publicly available DNA sequences. A new release is made every two months. GenBank is part of the International Nucleotide Sequence Database Collaboration, which is comprised of the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at the National Center for Biotechnology Information.

Swissprot (also known as UniProtKB/Swiss-Prot) is a curated protein sequence database (http://expasy.org/sprot/) maintained by the Swiss Institute of Bioinformatics (SIB). The database strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases.

These databases are updated weekly or monthly, and the sequences extend to those in the databases at the time of filing of this application. When finding sequences within the databases that have the desired homology, any method may be employed, depending on the match criteria. However, preferably the BLASTP program may be employed [BLAST and its derived programmes (e.g. BLASTP) are public domain software].

In another embodiment, instead of (or in addition to) the 85% homology referred to above in part (c), polypeptide sequences within the above databases are also included that have at least 85% of their amino acids in common (in terms of both amino acid identity, and position in the sequence) with a part of a sequence in (a) or (b) that 8 amino acids or more in length, and preferably from 8 amino acids in length up to one third of the length of the sequence in (a) or (b). In other words, for a sequence in (a) or (b) that is 30 amino acids long, a sequence in the database is also included if it shares 85% or more of its amino acids with any part of the sequence in (a) or (b) that is 8 amino acids long or greater, preferably from 8-10 amino acids. Similarly, if a sequence in (a) or (b) is 60 amino acids in length, a sequence in the database is also included if it shares 85% or more of its amino acids with any part of the sequence in (a) or (b) that is 8 amino acids long or greater, preferably from 8-20 amino acids. The matching amino acids need not be consecutive. For example, in the case of a 20 amino acid sequence in (a) or (b), where a corresponding database sequence shares 17 or more amino acids in the correct positions, it is included, even if these positions are not all consecutive.

Typically, the polypeptides in the composition are not complete (not full or not entire) arthropod saliva proteins. By complete (or full or entire) it is meant that the polypeptides do not contain all of the amino acid residues present in any of the naturally occurring arthropod saliva proteins.

Thus, the polypeptide is one that may comprise the whole of (or may comprise at least one 7 or more residue parts of) any of the above sequences. The polypeptide must also be immunogenic in a vertebrate. Typically this immunogenicity is such that the polypeptides cause the vertebrate to produce immune system cells capable of recognising at least one epitope from an arthropod saliva protein fraction,. Thus, where a polypeptide elicits a T-cell or B-cell response, it is immunogenic in a vertebrate. The polypeptide may alternatively be a T helper lymphocyte (Th) epitope, or may be a B lymphocyte epitope.

One method for determining whether a polypeptide possesses immunogenicity is set out in Experiment 2 below. However, the present invention is not limited to such methods, and the skilled person may select any known method for determining immunogenicity, as desired.

It is particularly preferred that the polypeptides of the polypeptide composition are selected independently from SEQ ID 1-6, 20, 28, 30-32 and 35, or sub-sequences from these sequences, the sub-sequences having 7 amino acids or more, or from polypeptide sequences having 85% homology or more with one of these sequences and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

Typically, but not exclusively, the polypeptide composition of the present invention, comprises 2 or more polypeptides, preferably from 2 to 10 polypeptides, or more preferably from 2-6 polypeptides. However, the composition may comprise a single polypeptide and further non-polypeptide components, if desired.

Generally, in the polypeptide composition according to the invention, the arthropod saliva protein fraction has a mass of40kDa or less, 30kDa or less, or more preferably 20kDa or less. The fraction may also have a mass of from 20-40kDa, from 20-30kDa, or from 10-20kDa.

In another embodiment of the invention, the polypeptide composition of the invention comprises SEQ ID 131, or comprises one or more sub-sequences of SEQ ID 131 having 7 amino acids or more, or comprises polypeptide sequences having 85% homology or more with one of these sequences and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

SEQ ID 131
FLKGSFPRFQMCVMLIGFFSSAKCLFYSVGKLVKVLLVMAVCCLLLCTAPTGADPL
MKFAFAFVLIALFAVFAVSQALPQPEQAAAGRVMCLLRLMSTLLVVLSIVGKLYSGY
RLLVLLVMTVCCLLLFIAPTGADPLPGQTQRTLALIGLLLCSVQSVTANDPVDALGA
CSGNLFGLLMTRLSKLFVLAFLCLALVVVVQSAPQYARGDVPTLLPAKVIPDKTAAY
VAYGGQETLVEHVEVLVRYFVVIALICPLIIVETLAVVLLTPALQAYIMDEHNLNRSN
IALGRIRPYPSAVKMPVLHSMLVNASLAEMVKESYQTHGADGRMVVRMLKFVRLL
PRVRALRALLETLLQHQGEQNNDVYLIRLAHETELQQALSSLNAGSGSCAEVFNAYL
PVHNKYIGVSRKIQYSMECLEAAEPKYLDGLKTLADETAQCSFAWLLYGIILRSNFL
VVQNLMALALSAVQLSLFIIAFPFISGFLSCFMWLKYGVLTEESTLILVNFIGSAL

In another embodiment of the invention the polypeptide composition of the invention comprises one or more sequences selected from SEQ ID 1-4, 6, 9, 10, 13, 14, 17-19, 21-25, 27, 29, 33, 34, 36-40, 43, and 44 or comprises one or more sub-sequences of these sequences having 7 amino acids or more, or comprises polypeptide sequences having 85% homology or more with one of these sequences and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

In a further embodiment, the polypeptide composition of the invention comprises SEQ ID 132, or comprises one or more sub-sequences of SEQ ID 132 having 7 amino acids or more, or comprises polypeptide sequences having 85% homology or more with one of these sequences and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

SEQ ID 132
HLTLFTVAVLLLAAAALLLLLPPAYSTTLTPPPLSYCHLFLTHTLARALSFSRSDCLKN
VFFALLLVVLVCCLVSVQGNEIKLLVLLICLFFYHTHCTTAYLWLAMGVNDYQALL
GLCCPWIDLAAADLPMRRHAKADGASAITKIVLELTPEQAAAVTLFIFLVCCQIPLFGI
MSSDSADPFYWIRVILAMYCVIKGKTGGYCNSEGLCTCRAEDLHFLLKPIINKDNAE
DPRTELIGCGSVLFHLAANRLSLQLEEFAVCKRSMLVAFATLSVALVVVVAIPANFN
YGGGGGYFINGTGQIYEKLPAYLSEVSARVNVLQVSLQHDLPNLQEMKLAKVALVT
ISLWFMAWTPYLVINFTGIFYTCFLGTSSLAGFKNAVDYDELLKAGVLEVLGFVEDN
GELVFQELLGVLKMVDPDGDKLTPTVVVVLLCLTFVADALTIQELRAQIAQQRIQQR
YGVTVATTSLSDYGLIELKEHCLECCQKDTEADSKLKVYPAAVLEVTYICFILHGVSE
IIPQQQKKTMKFLLLVASVLCLVLILLLYLDAADLRRALHQYQLLAAQGDRHLPQQI
VKFVVLKGETHKALKLKDGGHYLVEFKSIYMKFYRLISTLLVVVVIAPRHQCSPFFF
QYNRPYLNYVPDVSALEQDIIEVDPETKEMLKHLDFNNIVVQLEYAQVTKMLGNGR
LEAMCFDGVKRLCHIRGKLKLFLTLLSTLSVAMVFALPAHHHSRGELEEARLVAEEL
EERQQELDYLKRYLVGRLQAVSYFLTVCLLALVQSETVQDAMTNANLVGLTISLAY
AIFFLLYTPPTGRSSGLLCCCLAVLFFASPLTMLAHVIRLLLAMVLLPLLLLESVVPYA
AAEKVW

In a still further embodiment, the polypeptide composition of the invention comprises SEQ ID 133, or comprises one or more sub-sequences of SEQ ID 133 having 7 amino acids or more, or comprises polypeptide sequences having 85% homology or more with one of these sequences and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

SEQ ID 133
HLTLFTVAVLLLAAAALLLLLPPAYSTTLTPPPLSYCHLFLTHTLARALSFSRSDCLKN
VFFALLLVVLVCCLVSVQGNEIKLLVLLICLFFYHTHCTTAYLWLAMGVFLKGSFPR
FQMCVMLIGFFSSAKCLNDYQALLGLCCPWIDLAAADLPMRRHAKA

In a yet further embodiment, the polypeptide composition of the invention comprises SEQ ID 134, or comprises one or more sub-sequences of SEQ ID 134 having 7 amino acids or more, or comprises polypeptide sequences having 85% homology or more with one of these sequences and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

SEQ ID 134
LLPAKVIPDKTAAYVAYGGQETLVEHVEVLVVLLTPALQAYIMDEHNLNRSNIALGR
IRPYPSAVKMPVLHSMLVNASLAEMVKESYQTHGADGRMVVRMLKFVRLLPRVRA
LRALLETLLQHQGEQNNDVYLIRLAHETELQQALSSLNAGSGSCAEVFNAYLPVHN
KYIGVSRKIQYSMECLEAAEPKYLDGLKTLADETAQC

Importantly, in all embodiments of the invention, any combination of the listed sequences may be employed, either alone or in combination. Particularly preferred polypeptides are those having sequences of SEQ ID 1-6, and/or those having sequences of SEQ ID 20, 28, 30-32 and 35. It is especially preferred that any one or more of these are present, and particularly preferred that all of SEQ ID 1-6 and/or all of SEQ ID 20, 28, 30-32 and 35 are present.

The present inventors have found that the above sequences comprise an epitope, or a plurality of epitopes, which may afford protection against arthropod borne diseases for a wide variety of vertebrates in a population. Arthropod bites induce in the host an immune responses to the saliva components that is characterised by a Th2 phenotype (i.e. downregulation of IFN-γ production and upregulation of IL-4 production) (Mbow et al, 1998). This immune response, together with the antiheamostatic effect of many of these salivary molecules, has been found to facilitate and enhance parasite transmission in general (Dhar and Kumar, 2003) and to promote Leishmania infection in particular (Kamhawi et al., 2000). In contrast, an increase in the cellular immune response characterised by increased production of IFN-γ and IL-12, both Th1 type cytokines, at the site of infection (i.e. the bite site) has been shown to induce protection against *Leishmania major* infection by the bite of infected sandflies (Kamhawi et al., 2000).

Without being bound by theory, it is believed that immunisation with salivary proteins leading to the activation of a Th1 type response will lead to the rapid recognition of salivary antigens at the bite site by cells of the immune system (such as activated cytotoxic T cells (CTLs) and T helper type 1 cells) and the production of IFN-γ. This cytokine (1) stimulates both T and NK cells to produce more IFN-γ, (2) promotes the microbicidal activity of macrophages, (3) induces isotype switching to and increased production of IgG2a by B cells and, (4) the production of multiple cytokines (e.g. TNF-α, interleukin (IL) 12 and IL-18) which combined trigger a cascade of immune reactions that lead to the death of intracellular parasites. References detailing this are set out below.

Mbow ML, Bleyenberg JA, Hall LR & Titus RG. 1998. *Phlebotomus papatasi* sandfly salivary gland lysate down-regulates a Th1, but up-regulates a Th2, response in mice infected with Leishmania major. J. Immunol; 161: 5571–5577.

Dhar, R., Kumar, N., 2003. Role of mosquito salivary glands. Cur. Sci. 85, 1308–1313.
Kamhawi, S., Belkaid, Y., Modi, G., Rowton, E., Sacks, D., 2000. Protection against cutaneous leishmaniasis resulting from bites of uninfected sand flies. Science 290, 1351–1354

Malaguarnera L, Musumeci S. 2002. The immune response to *Plasmodium falciparum* malaria. Lancet Infect Dis. Aug;2(8):472-8

As discussed above, the sequences have been identified after analysis of saliva sequences in *Anopheles gambiae*. It will be apparent to the skilled person that the invention extends not only to the sequences and their epitopes, but also to larger sequences in arthropod saliva proteins containing these sequences, and to sequences that are homologous to these sequences and so have immunogenic activity. Thus, sequences with some homology to the consensus sequences are also within the scope of the invention. Such homology allows substitution of, for example, up to 3 amino acids in an 8-mer epitope (62.5% homology) or in a 9–mer, 10-mer, or 11-mer epitope. It is preferred that no more than 10 such substitutions are identifiable in a sequence of the invention corresponding to the full sequences of SEQ ID 1-44 (66.6% homology for a 30-mer). Such substitutions are preferably conservative substitutions in line with known substitution schemes.

The invention also provides a polypeptide comprising one or more sequences defined by the following sequences SEQ ID 1-44, or the following amino acid residues of an arthropod saliva protein, or comprising one or more sub-sequences from these sequences, the sub-sequences having 7 amino acids or more, or comprising polypeptide sequences having 85% homology or more with one of the sequences and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

SEQ ID 1	HLTLFTVAVLLLAAAALLLLLPPAYSTTLTPP
SEQ ID 2	PLSYCHLFLTHTLARALSFSRSDCL
SEQ ID 3	KNVFFALLLVVLVCCLVSVQGNEI
SEQ ID 4	KLLVLLICLFFYHTHCTTAYLWLAMGV
SEQ ID 5	FLKGSFPRFQMCVMLIGFFSSAKCL
SEQ ID 6	NDYQALLGLCCPWIDLAAADLPMRRHAKA
SEQ ID 7	FYSVGKLVKVLLVMAVCCLLLCTAPTGADPL
SEQ ID 8	MKFAFAFVLIALFAVFAVSQALPQPEQAAA
SEQ ID 9	DGASAITKIVLELTPEQAAAV
SEQ ID 10	TLFIFLVCCQIPLFGIMSSDSADPFYWIRVILA
SEQ ID 11	GRVMCLLRLMSTLLVVLSIVGK
SEQ ID 12	LYSGYRLLVLLVMTVCCLLLFIAPTGADPLPGQTQRTL
SEQ ID 13	MYCVIKGKTGGYCNSEGLCTCRAEDLHFLLKPIINKD
SEQ ID 14	NAEDPRTELIGCGSVLFHLAANRLSLQLEEFAVCKR
SEQ ID 15	ALIGLLLCSVQSVTANDPVDALGACSGNLFGLLMTRL
SEQ ID 16	SKLFVLAFLCLALVVVVQSAPQYARGDVPT

SEQ ID 17	SMLVAFATLSVALVVVVAIPANFNYGGGGGYFINGTGQ
SEQ ID 18	IYEKLPAYLSEVSARVNVLQVSLQHDLPNLQ
SEQ ID 19	EMKLAKVALVTISLWFMAWTPYLVINFTGI
SEQ ID 20	LLPAKVIPDKTAAYVAYGGQETLVEHVEVLV
SEQ ID 21	FYTCFLGTSSLAGFKNAVDYDELLKAG
SEQ ID 22	VLEVLGFVEDNGELVFQELLGVLKMVDPDGD
SEQ ID 23	KLTPTVVVVLLCLTFVADALTIQELRAQIAQQRIQQRYGVTVATT
SEQ ID 24	SLSDYGLIELKEHCLECCQKDTEADSKLKVYPAAVLEV
SEQ ID 25	TYICFILHGVSEIIPQQQKKTMKFLLLVASVLCLVLI
SEQ ID 26	RYFVVIALICPLIIVETLAV
SEQ ID 27	LLLYLDAADLRRALHQYQLLAAQGDRHLPQQIVKFV
SEQ ID 28	VLLTPALQAYIMDEHNLNRSNIALGRIRPYPSAVKMP
SEQ ID 29	VLKGETHKALKLKDGGHYLVEFKSIYM
SEQ ID 30	VLHSMLVNASLAEMVKESYQTHGADGRMVVRMLKFVRLLP
SEQ ID 31	RVRALRALLETLLQHQGEQNNDVYLIRLAHET
SEQ ID 32	ELQQALSSLNAGSGSCAEVFNAYLPVHNKYIGVSRKI
SEQ ID 33	KFYRLISTLLVVVVIAPRHQCSPFFFQYNRPYL
SEQ ID 34	NYVPDVSALEQDIIEVDPETKEMLKHLDFNNIVVQL
SEQ ID 35	QYSMECLEAAEPKYLDGLKTLADETAQC
SEQ ID 36	EYAQVTKMLGNGRLEAMCFDGVKRLCHIRGKL
SEQ ID 37	KLFLTLLSTLSVAMVFALPAHHHSRG
SEQ ID 38	ELEEARLVAEELEERQQELDYLKRYLVGRLQAV
SEQ ID 39	SYFLTVCLLALVQSETVQD
SEQ ID 40	AMTNANLVGLTISLAYAIFFLLYTPPTGRSS
SEQ ID 41	SFAWLLYGIILRSNFLVVQNLMALALSAVQLSLFII
SEQ ID 42	AFPFISGFLSCFMWLKYGVLTEESTLILVNFIGSAL
SEQ ID 43	GLLCCCLAVLFFASPLTMLAHVIR
SEQ ID 44	LLLAMVLLPLLLLESVVPYAAAEKVW

residues 2-33 of >gi|18389913|gb|AAL68793.1|AF457563_1 hypothetical protein 16 [Anopheles gambiae]

residues 2-26 of >gi|18389909|gb|AAL68791.1|AF457561_1 hypothetical protein 14 [Anopheles gambiae]

residues 2-25 of >gi|18389907|gb|AAL68790.1|AF457560_1 hypothetical protein 13 [Anopheles gambiae]

residues 10-36 of >gi|18389903|gb|AAL68788.1|AF457558_1 hypothetical protein 11 [Anopheles gambiae]

residues 2-26 of >gi|62546227|gb|AAX86005.1| hyp3.5 precursor [Anopheles gambiae]

residues 14-42 of gi|18389899|gb|AAL68786.1|AF457556_1 salivary gland 7-like protein [Anopheles gambiae]

residues 3-33 of >gi|18389911|gb|AAL68792.1|AF457562_1 hypothetical protein 15 [Anopheles gambiae]

residues 1-30 of >gi|62546225|gb|AAX86004.1| hyp6.3 precursor [Anopheles gambiae]

residues 34-54 of >gi|62546225|gb|AAX86004.1| hyp6.3 precursor [Anopheles gambiae]

residues 38-70 of >gi|17026153|emb|CAD12038.1| Sec61 protein [Anopheles gambiae]

residues 2-23 of >gi|62546223|gb|AAX86003.1| hyp6.2 precursor [Anopheles gambiae]

residues 17-54 of >gi|18389915|gb|AAL68794.1|AF457564_1 hypothetical protein 17 [Anopheles gambiae]

residues 57-93 of >gi|87080391|gb|ABD18596.1| defensin [Anopheles gambiae]

residues 22-57 of >gi|18389901|gb|AAL68787.1|AF457557_1 hypothetical protein 10 [Anopheles gambiae]

residues 7-43 of >gi|18389905|gb|AAL68789.1|AF457559_1 hypothetical protein 12 [Anopheles gambiae]

residues 3-32 of >gi|4127344|emb|CAA76832.1| cE5 protein [Anopheles gambiae]

residues 3-40 of >gi|4210617|emb|CAA10259.1| SG2 protein [Anopheles gambiae]

residues 91-121 of >gi|4127309|emb|CAA76820.1| hypothetical protein [Anopheles gambiae]

residues 65-94 of >gi|4375824|emb|CAA76825.1| opsin [Anopheles gambiae]

residues 41-71 of >gi|62546233|gb|AAX86008.1| unknown [Anopheles gambiae]

residues 117-143 of >gi|3378531|emb|CAA03872.1| D7r2 protein [Anopheles gambiae]

residues 63-93 of >gi|3378529|emb|CAA03871.1| D7r3 protein [Anopheles gambiae]

residues 23-67 of >gi|18389893|gb|AAL68783.1|AF457553_1 mucin-like protein [Anopheles gambiae]

residues 43-80 of >gi|18389881|gb|AAL68777.1|AF457547_1 selenoprotein [Anopheles gambiae]

residues 6-42 of >gi|18389879|gb|AAL68776.1|AF457546_1 30 kDa protein [Anopheles gambiae]

residues 4-23 of >gi|18378603|gb|AAL68639.1|AF458073_1 D7-related 5 protein [Anopheles gambiae]

residues 20-55 of >gi|18389897|gb|AAL68785.1|AF457555_1 salivary gland 1-like 4 protein [Anopheles gambiae]

residues 59-95 of >gi|18389883|gb|AAL68778.1|AF457548_1 antigen 5-related 1 protein [Anopheles gambiae]

residues 158-184 of >gi|83016748|dbj|BAE53441.1| DsRed [synthetic construct]

residues 37-76 of >gi|18389895|gb|AAL68784.1|AF457554_1 salivary gland 1-like 3 protein [Anopheles gambiae]

residues 191-222 of >gi|18389895|gb|AAL68784.1|AF457554_1 salivary gland 1-like 3 protein [Anopheles gambiae]

residues 113-149 of >gi|18389891|gb|AAL68782.1|AF457552_1 D7 protein long form [Anopheles gambiae]

residues 3-35 of >gi|13537676|emb|CAC35527.1| gSG9 protein [Anopheles gambiae]

residues 83-118 of sp|Q9U9L1|RS17_ANOGA 40S ribosomal protein S17 and gb|AAD47077.1|AF164153_1 ribosomal protein S17 [Anopheles gambiae]

residues 113-140 of >emb|CAC35523.1| gSG7 protein [Anopheles gambiae]

residues 34-65 of >gb|AAD47075.1|AF164151_1 translation initiation factor 4C (1A) [Anopheles gambiae]

residues 2-27 of >emb|CAC35519.1| gSG2-like protein [Anopheles gambiae]

residues 108-140 of >emb|CAC35451.1| hypothetical protein [Anopheles gambiae] and >gb|EAU75730.1| ENSANGP00000031975 [Anopheles gambiae str. PEST]

residues 8-26 of >emb|CAC35524.1| D7r4 protein [Anopheles gambiae] and >gb|AAK84945.1| D7-related 4 protein [Anopheles gambiae] residues 72-102 of >ref|XP_001230998.1| ENSANGP00000014906 [Anopheles gambiae str. PEST] and gb|EAU76798.1| ENSANGP00000014906 [Anopheles gambiae str. PEST]

residues 176-211 of >ref|XP_316361.2| ENSANGP00000012984 [Anopheles gambiae str. PEST] and >gb|EAA10852.2| ENSANGP00000012984 [Anopheles gambiae str. PEST]

residues 43-78 of >ref|XP_314140.3| ENSANGP00000015780 [Anopheles gambiae str. PEST] and gb|EAA09398.3| ENSANGP00000015780 [Anopheles gambiae str. PEST]

residues 128-151 of >ref|XP_314140.3| ENSANGP00000015780 [Anopheles gambiae str. PEST] and gb|EAA09398.3| ENSANGP00000015780 [Anopheles gambiae str. PEST]

residues 7-32 of >emb|CAC35522.1| gSG6 protein [Anopheles gambiae]

Preferably, the polypeptide is not a complete arthropod saliva protein.

The sequence numbering referred to in the present invention is defined according to well-recognised principles. Thus, the numbering begins at 1 from the recognised translation initiation codon (ATG). This corresponds to a Methionine (M), for the segment of the genome coding for the protein of interest. In other words, it begins at 1 in respect of the Methionine shown as the first amino acid in the protein sequence of interest as used and defined by the databases in which the sequences have been set forth (i.e. GenBank, SwissProt, etc.).

The present invention will be described in more detail by way of example only with reference to the following Figures, in which:

Figures 1 and 9 show IEF gels of anopheles gambiae salivary glands (Comassie Blue stain), as produced in accordance with the protocols set out in the Examples;

Figure 2 shows an SDS-PAGE of the IEF gel of Figure 1 (silver stain) – the red rectangle indicates the location of one of the targeted proteins (< 30 kDa)

Figures 3A-3I show data on the effect of the vaccine on the fecundity of mosquitoes, as follows:

3A : Per cent that fed
3B : Mean no eggs produced
3C : Mean no eggs laid
3D : Per cent hatch rate
3E : Mean no larvae
3F : Mean no pupae
3G : Per cent pupation
3H : Per cent emergence
3I : Mean no adults

Figure 4 shows the data on the per cent survival of mosquitoes after feeding on immunized mice.

Figure 5 shows IFN-gamma production following 96 hour stimulation *in vitro* with the antigens in Experiment 3.

Figure 6 shows the total Ig response in sera to the antigens according to Experiment 3.

Figure 7 shows total Ig response in sera to the AGS-mix at day 21 in accordance with Experiment 4.

Figure 8 shows the AGS-mix immunised group of Experiment 4 displaying an increased survival rate as compared with those in the control NRP-mix immunised group.

The polypeptide described above typically comprises one or more (preferably two or more) epitopes. These epitopes are preferably T cell epitopes, such as cytotoxic T lymphocyte (CTL) epitopes, but may also contain B cell epitopes. Generally the polypeptide is immunogenic to an arthropod saliva protein, and preferably to a plurality of such proteins. In the present context, a polypeptide immunogenic to an arthropod saliva protein is understood to mean a polypeptide that is part of an arthropod saliva protein and that elicits an immune system response. One method for determining whether a polypeptide possesses such immunogenicity is set out in Experiment 2 below. However, the present invention is not limited to such methods, and the skilled person may select any known method for determining immunogenicity, as desired.

In the present invention, the polypeptide composition comprises one or more sequences as described above. Typically, two, three, four, five or more such sequences may be present in the polypeptide, if desired. The more such epitopes are present, the greater the breadth of protection afforded within a population of humans and/or animals individuals with differing HLAs or MHCs. This is particularly so if the epitopes included are from the saliva of a plurality of differing arthropods or shared by salivary proteins of different arthropod species, and can thus offer protection against diseases carried by a plurality of different arthropods. Typically the polypeptide composition comprises 10 polypeptides or fewer, preferably 6 polypeptides or fewer, and typically from 2-10 polypeptides, and more preferably from 2-6 polypeptides.

The polypeptide composition according to the present invention may also comprise one or more further sequences that are not epitopes, if desired. Typically the further sequences are from one or more arthropod saliva proteins, preferably selected from the sequences of SEQ ID 45-85 or sub-sequences of these. These sequences may be situated between two or more of the sequences (the epitopes) described above, or may be situated at one or both ends of the polypeptide. The presence of such further sequences should not affect the function of the polypeptide, provided that the polypeptide as a whole does not become too large, interfering with the presentation of the epitopes in the vertebrate's immune system.

In the most preferred embodiments, the further sequences from the above-mentioned proteins are ones that are (or are within) the following sequences:

SEQ ID 45 - >gi|18389913|gb|AAL68793.1|AF457563_1 hypothetical protein 16 [Anopheles gambiae]
MHLTLFTVAVLLLAAAALLLLLPPAYSTTLTPPAPPRLSHLGITIGRI

SEQ ID 46 - >gi|18389909|gb|AAL68791.1|AF457561_1 hypothetical protein 14 [Anopheles gambiae]
MPLSYCHLFLTHTLARALSFSRSDCLKFSEKRLLFSGSKTFPTTLL

SEQ ID 47 - >gi|18389907|gb|AAL68790.1|AF457560_1 hypothetical protein 13 [Anopheles gambiae]
MKNVFFALLLVVLVCCLVSVQGNEIIQNVVKRSIPLRQLILQHNALDDSNSDSGSQ

SEQ ID 48 - >gi|18389903|gb|AAL68788.1|AF457558_1 hypothetical protein 11 [Anopheles gambiae]
MCIFFQAGIKLLVLLICLFFYHTHCTTAYLWLAMGVEAKSIKARGTAHSKSRTSTN

SEQ ID 49 - >gi|62546227|gb|AAX86005.1| hyp3.5 precursor [Anopheles gambiae]
MFLKGSFPRFQMCVMLIGFFSSAKCLMCFADWEGMLLMTMEVFDFQLIVFTPVLKR
S

SEQ ID 50 - >gi|18389899|gb|AAL68786.1|AF457556_1 salivary gland 7-like protein [Anopheles gambiae]
MAGESQKNARSKQNDYQALLGLCCPWIDLAAADLPMRRHAKAREAINFLLQAHEA
GPNEEPSLPA

SEQ ID 51 - >gi|18389911|gb|AAL68792.1|AF457562_1 hypothetical protein 15 [Anopheles gambiae]
MKFYSVGKLVKVLLVMAVCCLLLCTAPTGADPLPGRDRNTIANKSKDKKASAPKHS
LGTGARMALTGGGVLGGVLTNM

SEQ ID 52 - >gi|62546225|gb|AAX86004.1| hyp6.3 precursor [Anopheles gambiae]
MKFAFAFVLIALFAVFAVSQALPQPEQAAASSNDGASAITKIVLELTPEQAAAVQKM
GGRGFWPIMMKSVKKIMAIGCDLIDC

SEQ ID 53 - >gi|17026153|emb|CAD12038.1| Sec61 protein [Anopheles gambiae]
MGIKFLEIIKPFCGILPEIAKPERKIQFREKVLWTAITLFIFLVCCQIPLFGIMSSDSADPF
YWIRVILASNRGTLM

SEQ ID 54 - >gi|62546223|gb|AAX86003.1| hyp6.2 precursor [Anopheles gambiae]
MGRVMCLLRLMSTLLVVLSIVGKKTNAAPQVTEAPGNVGSTYSPMADIGRLATGAT
KLFGQFWNTGTRFGTELSRRTFDFLRVKK

SEQ ID 55 - >gi|18389915|gb|AAL68794.1|AF457564_1 hypothetical protein 17 [Anopheles gambiae]
MAGDIQLFSTRETTMKLYSGYRLLVLLVMTVCCLLLFIAPTGADPLPGQTQRTLGYR
GNDKRATPPMHSLGSGARMAMTGGGILGGIFSAL

SEQ ID 56 - >gi|87080391|gb|ABD18596.1| defensin [Anopheles gambiae]
MDQCSVPRLCIIIMKSFIAAAVIALICAIAVSGTTVTLQSTCKLFTADVVSSITCKMYC
VIKGKTGGYCNSEGLCTCRAEDLHFLLKPIINKD

SEQ ID 57 - >gi|18389901|gb|AAL68787.1|AF457557_1 hypothetical protein 10 [Anopheles gambiae]
MRFLSVLTVGLLVWVGVFATVNAEDPRTELIGCGSVLFHLAANRLSLQLEEFAVCK
RSNPGYDCSDSIHRAISDLQQGLFDLNHCTKDIR

SEQ ID 58 - >gi|18389905|gb|AAL68789.1|AF457559_1 hypothetical protein 12 [Anopheles gambiae]
MRFCCVALIGLLLCSVQSVTANDPVDALGACSGNLFGLLMTRLQQMVEDFTACRQE
ATANDPQHDRSDSIQRAKVDLQQQLVNYSYCTKNIQ

SEQ ID 59 - >gi|4127344|emb|CAA76832.1| cE5 protein [Anopheles gambiae]
MASKLFVLAFLCLALVVVVQSAPQYARGDVPTYDEEDFDEESLKPHSSSPSDDGEEE
FDPSLLEEHADAPTARDPGRNPEFLRNSNTDEQASAPAASSSDS

SEQ ID 60 - >gi|4210617|emb|CAA10259.1| SG2 protein [Anopheles gambiae]
MKSMLVAFATLSVALVVVVAIPANFNYGGGGGYFINGTGQSFNFSGESNGTSIPGLP
DFGSFLPNLGNLTQQFGGSSGAFPQFSIPSWTNFTDAFTSILPFFGNGQGGGFPFFG

SEQ ID 61 - >gi|4127309|emb|CAA76820.1| hypothetical protein [Anopheles gambiae]
MTPLIATLAACALTLSIVHSRGLPESSDKLEACGQHYGXLLKASTTWNEKECNGSTK
LAACVVSEHEQAYRELKQRCQEAHDERTAKVNAIYEKLPAYLSEVSARVNVLQVSL
QHDLPNLQE

SEQ ID 62 - >gi|4375824|emb|CAA76825.1| opsin [Anopheles gambiae]
PDVAEPLVHHHLRHLRVLAAAADHHLLVHLHPEGCVRSREEHARAGQEGNVASLR
TQEAQNTSTEMKLAKVALVTISLWFMAWTPYLVINFTGIFKAAPISPLATIRGSLFAK
ANAVYNPIVYG

SEQ ID 63 - >gi|62546233|gb|AAX86008.1| unknown [Anopheles gambiae]
MATTWIPTSVHGPYPPHMVPGGVDSDGAQIFVGRAHHAGDLLPAKVIPDKTAAYVA
YGGQETLVEHVEVLVHKQLIWDTASAGQVPLGAVVGGHTSDGEILYVGRAYHEGS
QTIGKVQCSHNCIYIPYGGAEVSVPTYEVLCER

SEQ ID 64 - >gi|3378531|emb|CAA03872.1| D7r2 protein [Anopheles gambiae]
MFKKLLLSVGLVWCLISLGQARKESTVEECEKNIGDSLKDRVCELRQYTPVSSDDM
DKHMQCVLEVVGFVDGNGEVKESVLLELLQRVDSGVNHAANMKKCVTEASTSGSD
KKANTFYTCFLGTSSLAGFKNAVDYDELLKAGKMQTSDP

SEQ ID 65 - >gi|3378529|emb|CAA03871.1| D7r3 protein [Anopheles gambiae]
MFGKLLPCAILLWCLFSLGQARQEETVEECERNIPASLKERVCELRQYTPVQGKDMD
SHMQCVLEVLGFVEDNGELVFQELLGVLKMVDPDGDHAGSMKKCNGEAEKVDTSS
KANTFYTCFLGTSSAQAFKYAVDYVXAXRAGKLDMGTTFNAGQV

SEQ ID 66 - >gi|18389893|gb|AAL68783.1|AF457553_1 mucin-like protein [Anopheles gambiae]

AGGFSLFEALKQTTTRGEMFRRKLTPTVVVVLLCLTFVADALTIQELRAQIAQQRIQQ
RYGVTVATTSAATTTAATTSAATTSEATTTAAASTTQASDSDNTTTTAEATTTTEAQ
TTSSSDNSTTTEAAATTTAASETTADSSSTGTTSVEAGLRAQYRDQVRQQAIERALA
RAAAFG

SEQ ID 67 - >gi|18389881|gb|AAL68777.1|AF457547_1 selenoprotein [Anopheles gambiae]

MRLFAITCLLFSIVTVIGAEFSAEDCRELGLIKSQLFCSACSSLSDYGLIELKEHCLECC
QKDTEADSKLKVYPAAVLEVCTCKFGAYPQIQAFIKSDRPAKFPNLTIKYVRGLDPIV
KLMDEQGTVKETLSINKWNTDTVQEFFETRLAKVEDDDYIKTNRV

SEQ ID 68 - >gi|18389879|gb|AAL68776.1|AF457546_1 30 kDa protein [Anopheles gambiae]

MAGAITYICFILHGVSEIIPQQQKKTMKFLLLVASVLCLVLIVSARPADDTSDQESSTE
LSDDAGAEEGAEDAGSDAEADAGAADGEEGATDTESGAEGDDSEMDSAMKEGEEG
AGSDDAVSGADDETEESKDDAEEDSEEGGEEGGDSASGGEGGEKESPRNTYRQVHK
LLKKIMKVDTKD

SEQ ID 69 - >gi|18378603|gb|AAL68639.1|AF458073_1 D7-related 5 protein [Anopheles gambiae]

MEWRYFVVIALICPLIIVETLAVSDCVRHVSESARNTVCDVRQYRVTKGVEADRYVQ
CFMTALGFADESGSIQRSNVLTALDAVETHDGVYTDAVDVCLSKAKKLPGTERSGY
FFSCMLRTESALNFRDAVELQELRVASKWPEGERFDRSKVQQMMRELNSQLRC

SEQ ID 70 - >gi|18389897|gb|AAL68785.1|AF457555_1 salivary gland 1-like 4 protein [Anopheles gambiae]

GREAIETMRTEQRNHRQQLLLLYLDAADLRRALHQYQLLAAQGDRHLPQQIVKFVY
AAPRHENRRLENLLDLVRQLPARQDQRTLYQLLQPEIMKRPAQNQSTLAMLTALEM
GQVVEGNGELKKQQDAMYQLVLKRWMFLCLAGQYREIVQFATKHPRLFE

SEQ ID 71 - >gi|18389883|gb|AAL68778.1|AF457548_1 antigen 5-related 1 protein [Anopheles gambiae]
MAIWIVCATLLLAVLSVVSVGGQYCSSDLCPRGGPHVGCNPPSSSGGPTCQGKQKA
RKVLLTPALQAYIMDEHNLNRSNIALGRIRPYPSAVKMPTLTWDPELASLADANARS
CNYGHDRCRATKKFPYAGQNIAITQFFGYRFTEKDLIHKFVSSWWSEYLDARPEHVR
KYPSSYSG

SEQ ID 72 - >gi|83016748|dbj|BAE53441.1| DsRed [synthetic construct]
MKLASSENVITEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLKVTKGGPL
PFAWDILSPQFQYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVATVTQD
SSLQDGCFIYKVKFIGVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGETHKALK
LKDGGHYLVEFKSIYMAKKPVQLPGYYYVDAKLDITSHNEDYTIVEQYERTEGRHH
LFLRSRAPPPPPLT

SEQ ID 73 - >gi|18389895|gb|AAL68784.1|AF457554_1 salivary gland 1-like 3 protein [Anopheles gambiae]
MAGQRHLIEQAWQYGAQLQHELMLTSMESDRVQRALVLHSMLVNASLAEMVKES
YQTHGADGRMVVRMLKFVRLLPGADERVAVYKQLAELLKSNGQDGRFPAVIFSTD
VRQLEDRYKPDHAQYEGKVVERWLAELQAGTFHEVVEFARDYPEYFARVEEPLYE
TLKQQWSAEGLDRMVSFPNALPVGVQRVRALRALLETLLQHQGEQNNDVYLIRLA
HETGRVEATVGQADAAVRQALDDVKKLFEQFKYQRGFPDYEALYKLFKGL

SEQ ID 74 - >gi|18389891|gb|AAL68782.1|AF457552_1 D7 protein long form [Anopheles gambiae]
MIVPRVLLFILLELFVQATQAFKALDPEEAWYVYERCHEDHLPSGPNRETYLKTWKF
WKLEPNDAVTHCYVKCTLAGLQMYDEKTNTFKPETVPVQHEAYKSFTEVESSKVN
ELQQALSSLNAGSGSCAEVFNAYLPVHNKYIGVSRKIYHGTVDSVAKIYEAKPEIKK
QEESFFAYCAKKALGANGKEGYKKIRDYELADSAEFRNAMDCVFRGFRYMDDSGL
KVDEVVRDFNLINKSDLEPEVRSVLASCTGTHAYDYYSCLLNSSVKEDFRNAFYFHE
LRSANYGYLAMGKVYEGPEKVKEELKKLNY

SEQ ID 75 - >emb|CAC35527.1| gSG9 protein [Anopheles gambiae]
MCKFYRLISTLLVVVVIAPRHQCSPFFFQYNRPYLSQPSSQLASTAANVVQRSNVTVA
LGNRINTDTALDDYGTRV

SEQ ID 76 - >sp|Q9U9L1|RS17_ANOGA 40S ribosomal protein S17
MGRVRTKTIKKASKVIIEKYYTRLTMDFDTNKRIVEEVAIIPTKPLRNKIAGFVTHLM
KRLRHSQVRGISIKLQEEERERRDNYVPDVSALEQDIIEVDPETKEMLKHLDFNNIVV
QLTNPTAPGYSNRRN

SEQ ID 77 - >emb|CAC35523.1| gSG7 protein [Anopheles gambiae]
MHAKPAFVLIALGVICLLQTTPTSASTNHVQQLMKVFRSMTQNFDYTKKPSYLQRA
KYGVQNQLRNPLVQKAGNLPKSAKLSDGCLKQMVARVTDLEASFYASFSYNCHDH
DQYSMECLEAAEPKYLDGLKTLADETAQCMRDQQ

SEQ ID 78 - >gb|AAD47075.1|AF164151_1 translation initiation factor 4C (1A) [Anopheles gambiae]
MPKNKGKGGKNRRRGKNENESEKRELIFKEDEQEYAQVTKMLGNGRLEAMCFDGV
KRLCHIRGKLRKKVWINQGDIILIGLRDYQDSKADVILKYTPDEARNLKTYGEFPESV
RINETVTFVENDMDDDIEFGDDYSSSEEGDAIDAI

SEQ ID 79 - >emb|CAC35519.1| gSG2-like protein [Anopheles gambiae]
MKLFLTLLSTLSVAMVFALPAHHHSRGGDGSSANSTGNSDNNSAGVPDFGFNSQSN
VPGFGNGQQPGQQQQGQQGQGFPFFGQGQSGFPSFGNRLQPFFGQNQQGQDGDAQ
QGRGVPFFGQGGGQGGIPSFGSGQQNGGVPFLGNGQGQSGFPSFGNGQQGGNFPFF
G

SEQ ID 80 - >emb|CAC35451.1| hypothetical protein [Anopheles gambiae]
MKLYAFALVLCVGLAVGAEVDSVPEVPSDLQQQLDELQLADKPEAPVDDAEQPLPP
NGDELPEDAPEPVPEDGSPDEEHLEEEQEEEAEADEEEADESESEESEESDELEEARL
VAEELEERQQELDYLKRYLVGRLQAVAILDRRVRPAVIRRPWIRRPWIRRPG

SEQ ID 81 > emb|CAC35524.1| D7r4 protein [Anopheles gambiae]
MIRQVIISYFLTVCLLALVQSETVQDCENKLPPSLKSRLCEIRRYEIIEGPEMDKHIHCV
MRALDFVYEDGRGDYHKLYDPLNIIELDKRHDVNLEKCIGECVQVPTSERAHVFYK
CLLKSTTGRTFKKVFDLMELKKAGKVPQHQRYTAEFVQIMKDYDKALNC

SEQ ID 82 - >ref|XP_001230998.1| ENSANGP00000014906 [Anopheles gambiae str. PEST]
MEAISEALQPYKEQVGMAAGILTVGQMFSGCFVCNDIRKKGTTDGFSAMPFVGGCG
LTVLFLQHGMLMNDSAMTNANLVGLTISLAYAIFFLLYTPPTGRSSYWRQVGGTALF
TITLLGYVKVENPSVVEDRFGMIITVLMLALIGQPLFGLPDIIRRKSTEGLPFAMILSGT
IVGLSWLLYGVILNNVFVVCQNLAAVTLSGIQLALFAIYPSKAAPPSKKRE

SEQ ID 83 - >ref|XP_316361.2| ENSANGP00000012984 [Anopheles gambiae str. PEST]
MESIAVALQPYKDTVGLTAAIVTVVQFFSGVLALNAIRRQGNTRGFSALPFLGGTVF
CLLNIQFGQMLRDDGMIRVNFIGLALNLLYVCGFYLYTEGPAKTAVWGQIGLAGAL
TAGVLSYVQYEDPQLVEFRFGLILTGLLWTLVGMPLLGLGDILKKKSTEGLPFPIIFLG
AVVSFAWLLYGIILRSNFLVVQNLMALALSAVQLSLFIIFPSGAAKPPPTPAKKRN

SEQ ID 84 - >ref|XP_314140.3| ENSANGP00000015780 [Anopheles gambiae str. PEST]
MDGIMSKGSLASLATVATVLQFLTGTVICNRYIRKKSTGDTSAFPFISGFLSCFMWLK
YGVLTEESTLILVNFIGSALFFSYTVVFFIFCVNKREVIRQMMVISCIILSATLYTLFETD
DEKSIRVIGLLCCCLAVLFFASPLTMLAHVIRTQNTDSLPFPIIMASFFVCLLWTAYGV
LIGDRFIQIPNLLGGILAGIQLTLYVIYPKKKASFSGGPRYSPLVSENPIL

SEQ ID 85 - >emb|CAC35522.1| gSG6 protein [Anopheles gambiae]
MAIRVELLLAMVLLPLLLLESVVPYAAAEKVWVDRDKVYCGHLDCTRVATFKGER
FCTLCDTRHFCECKETREPLPYMYACPGTEPCQSSDRLGSCSKSMHDVLCDRIDQAF
LEQ

The peptides of the present invention, such as those of SEQ ID 1-44 and those within SEQ ID 131-134 described above preferably comprise one or more further amino acids at one or both of their termini in order to aid in their processing into vaccines. Typically, these further amino acids are the ones adjacent to each of the termini of SEQ ID 1-44, as shown in the larger proteins of SEQ ID 45-85 above (these larger proteins contain the sequences of SEQ ID 1-44). Preferably the number of further amino acids at each terminus is from 1-5, more preferably from 1-3, and most preferably 2 at each terminus. In each of these cases, if there are less than two further amino acids at that terminus of the sequences of SEQ ID 1-44, then the further amino acids include all of the remaining amino acid(s) at that terminus. Particularly preferred sequences of this type, corresponding to SEQ ID 1-44, are as follows:

SEQ ID 86	MHLTLFTVAVLLLAAAALLLLLPPAYSTTLTPPAP
SEQ ID 87	MPLSYCHLFLTHTLARALSFSRSDCLKF
SEQ ID 88	MKNVFFALLLVVLVCCLVSVQGNEIIQ
SEQ ID 89	GIKLLVLLICLFFYHTHCTTAYLWLAMGVEA
SEQ ID 90	MFLKGSFPRFQMCVMLIGFFSSAKCLMC
SEQ ID 91	KQNDYQALLGLCCPWIDLAAADLPMRRHAKARE
SEQ ID 92	MKFYSVGKLVKVLLVMAVCCLLLCTAPTGADPLPG
SEQ ID 93	MKFAFAFVLIALFAVFAVSQALPQPEQAAASS
SEQ ID 94	SNDGASAITKIVLELTPEQAAAVQK
SEQ ID 95	AITLFIFLVCCQIPLFGIMSSDSADPFYWIRVILASN
SEQ ID 96	MGRVMCLLRLMSTLLVVLSIVGKKT
SEQ ID 97	MKLYSGYRLLVLLVMTVCCLLLFIAPTGADPLPGQTQRTLGY
SEQ ID 98	CKMYCVIKGKTGGYCNSEGLCTCRAEDLHFLLKPIINKD
SEQ ID 99	TVNAEDPRTELIGCGSVLFHLAANRLSLQLEEFAVCKRSN
SEQ ID 100	CVALIGLLLCSVQSVTANDPVDALGACSGNLFGLLMTRLQQ
SEQ ID 101	MASKLFVLAFLCLALVVVVQSAPQYARGDVPTYD
SEQ ID 102	MKSMLVAFATLSVALVVVVAIPANFNYGGGGGYFINGTGQSF
SEQ ID 103	NAIYEKLPAYLSEVSARVNVLQVSLQHDLPNLQE
SEQ ID 104	STEMKLAKVALVTISLWFMAWTPYLVINFTGIFK
SEQ ID 105	GDLLPAKVIPDKTAAYVAYGGQETLVEHVEVLVHK
SEQ ID 106	NTFYTCFLGTSSLAGFKNAVDYDELLKAGKM
SEQ ID 107	QCVLEVLGFVEDNGELVFQELLGVLKMVDPDGDHA
SEQ ID 108	RRKLTPTVVVVLLCLTFVADALTIQELRAQIAQQRIQQRYGVTVATTSA
SEQ ID 109	CSSLSDYGLIELKEHCLECCQKDTEADSKLKVYPAAVLEVCT

SEQ ID 110	AITYICFILHGVSEIIPQQQKKTMKFLLLVASVLCLVLIVS
SEQ ID 111	EWRYFVVIALICPLIIVETLAVSD
SEQ ID 112	QLLLLYLDAADLRRALHQYQLLAAQGDRHLPQQIVKFVYA
SEQ ID 113	RKVLLTPALQAYIMDEHNLNRSNIALGRIRPYPSAVKMPTL
SEQ ID 114	DGVLKGETHKALKLKDGGHYLVEFKSIYMAK
SEQ ID 115	ALVLHSMLVNASLAEMVKESYQTHGADGRMVVRMLKFVRLLPGA
SEQ ID 116	VQRVRALRALLETLLQHQGEQNNDVYLIRLAHETGR
SEQ ID 117	VNELQQALSSLNAGSGSCAEVFNAYLPVHNKYIGVSRKIYH
SEQ ID 118	MCKFYRLISTLLVVVVIAPRHQCSPFFFQYNRPYLSQ
SEQ ID 119	RDNYVPDVSALEQDIIEVDPETKEMLKHLDFNNIVVQLTN
SEQ ID 120	HDQYSMECLEAAEPKYLDGLKTLADETAQCMR
SEQ ID 121	EQEYAQVTKMLGNGRLEAMCFDGVKRLCHIRGKLRK
SEQ ID 122	MKLFLTLLSTLSVAMVFALPAHHHSRGGD
SEQ ID 123	SDELEEARLVAEELEERQQELDYLKRYLVGRLQAVAI
SEQ ID 124	IISYFLTVCLLALVQSETVQDCE
SEQ ID 125	DSAMTNANLVGLTISLAYAIFFLLYTPPTGRSSYW
SEQ ID 126	VVSFAWLLYGIILRSNFLVVQNLMALALSAVQLSLFIIFP
SEQ ID 127	TSAFPFISGFLSCFMWLKYGVLTEESTLILVNFIGSALFF
SEQ ID 128	VIGLLCCCLAVLFFASPLTMLAHVIRTQ
SEQ ID 129	VELLLAMVLLPLLLLESVVPYAAAEKVWVD

Particularly preferred such peptides also include the following:

residues 1-35 of >gi|18389913|gb|AAL68793.1|AF457563_1 hypothetical protein 16 [Anopheles gambiae]
residues 1-28 of >gi|18389909|gb|AAL68791.1|AF457561_1 hypothetical protein 14 [Anopheles gambiae]
residues 1-27 of >gi|18389907|gb|AAL68790.1|AF457560_1 hypothetical protein 13 [Anopheles gambiae]
residues 8-38 of >gi|18389903|gb|AAL68788.1|AF457558_1 hypothetical protein 11 [Anopheles gambiae]
residues 1-28 of >gi|62546227|gb|AAX86005.1| hyp3.5 precursor [Anopheles gambiae]

residues 12-44 of gi|18389899|gb|AAL68786.1|AF457556_1 salivary gland 7-like protein [Anopheles gambiae]

residues 1-35 of >gi|18389911|gb|AAL68792.1|AF457562_1 hypothetical protein 15 [Anopheles gambiae]

residues 1-32 of >gi|62546225|gb|AAX86004.1| hyp6.3 precursor [Anopheles gambiae]

residues 32-56 of >gi|62546225|gb|AAX86004.1| hyp6.3 precursor [Anopheles gambiae]

residues 36-72 of >gi|17026153|emb|CAD12038.1| Sec61 protein [Anopheles gambiae]

residues 1-25 of >gi|62546223|gb|AAX86003.1| hyp6.2 precursor [Anopheles gambiae]

residues 15-56 of >gi|18389915|gb|AAL68794.1|AF457564_1 hypothetical protein 17 [Anopheles gambiae]

residues 55-93 of >gi|87080391|gb|ABD18596.1| defensin [Anopheles gambiae]

residues 20-59 of >gi|18389901|gb|AAL68787.1|AF457557_1 hypothetical protein 10 [Anopheles gambiae]

residues 5-45 of >gi|18389905|gb|AAL68789.1|AF457559_1 hypothetical protein 12 [Anopheles gambiae]

residues 1-34 of >gi|4127344|emb|CAA76832.1| cE5 protein [Anopheles gambiae]

residues 1-42 of >gi|4210617|emb|CAA10259.1| SG2 protein [Anopheles gambiae]

residues 89-122 of >gi|4127309|emb|CAA76820.1| hypothetical protein [Anopheles gambiae]

residues 63-96 of >gi|4375824|emb|CAA76825.1| opsin [Anopheles gambiae]

residues 39-73 of >gi|62546233|gb|AAX86008.1| unknown [Anopheles gambiae]

residues 115-145 of >gi|3378531|emb|CAA03872.1| D7r2 protein [Anopheles gambiae]

residues 61-95 of >gi|3378529|emb|CAA03871.1| D7r3 protein [Anopheles gambiae]

residues 21-69 of >gi|18389893|gb|AAL68783.1|AF457553_1 mucin-like protein [Anopheles gambiae]

residues 41-82 of >gi|18389881|gb|AAL68777.1|AF457547_1 selenoprotein [Anopheles gambiae]

residues 4-44 of >gi|18389879|gb|AAL68776.1|AF457546_1 30 kDa protein [Anopheles gambiae]

residues 2-25 of >gi|18378603|gb|AAL68639.1|AF458073_1 D7-related 5 protein [Anopheles gambiae] residues 18-57 of >gi|18389897|gb|AAL68785.1|AF457555_1 salivary gland 1-like 4 protein [Anopheles gambiae]

residues 57-97 of >gi|18389883|gb|AAL68778.1|AF457548_1 antigen 5-related 1 protein [Anopheles gambiae]

residues 156-186 of >gi|83016748|dbj|BAE53441.1| DsRed [synthetic construct]

residues 35-78 of >gi|18389895|gb|AAL68784.1|AF457554_1 salivary gland 1-like 3 protein [Anopheles gambiae]

residues 189-224 of >gi|18389895|gb|AAL68784.1|AF457554_1 salivary gland 1-like 3 protein [Anopheles gambiae]

residues 111-151 of >gi|18389891|gb|AAL68782.1|AF457552_1 D7 protein long form [Anopheles gambiae]

residues 1-37 of >emb|CAC35527.1| gSG9 protein [Anopheles gambiae]

residues 81-120 of >sp|Q9U9L1|RS17_ANOGA 40S ribosomal protein S17

residues 111 to 142 of >emb|CAC35523.1| gSG7 protein [Anopheles gambiae]

residues 32-67 of >gb|AAD47075.1|AF164151_1 translation initiation factor 4C (1A) [Anopheles gambiae]

residues 1-29 of >emb|CAC35519.1| gSG2-like protein [Anopheles gambiae]

residues 106-142 of >emb|CAC35451.1| hypothetical protein [Anopheles gambiae]

residues 6-28 of > emb|CAC35524.1| D7r4 protein [Anopheles gambiae]

residues 70-104 of >ref|XP_001230998.1| ENSANGP00000014906 [Anopheles gambiae str. PEST]

residues 174-213 of >ref|XP_316361.2| ENSANGP00000012984 [Anopheles gambiae str. PEST]

residues 41-80 of >ref|XP_314140.3| ENSANGP00000015780 [Anopheles gambiae str. PEST]

residues 126-153 of >ref|XP_314140.3| ENSANGP00000015780 [Anopheles gambiae str. PEST]

residues 5-34 of >emb|CAC35522.1| gSG6 protein [Anopheles gambiae]

In alternative embodiments of the present invention, the invention is directed to compositions comprising polypeptides which are homologous to those described above, in particular peptides that are homologous to any of SEQ ID 1-134. The homology referred to above in respect of these sequences is preferably 60%, 75%, 80%, 85%, 90%, 95% or substantially 100%.

The percent homology of a first polypeptide sequence to a second polypeptide sequence, as referred to in the context of the present invention, is defined as the number of amino acid residues in the second sequence that match in both position and identity to those in the first sequence, divided by the total number of amino acid residues in the second polypeptide (both first and second polypeptides must have the same number of amino acid residues) and multiplied by 100. In the present invention, it is preferred that the polypeptide homology to the defined sequences is 75% or more, 80% or more, 85% or more, 90% or more, 95% or more or 100% (or substantially 100%).

In the present invention, the arthropod borne disease is not especially limited, and the polypeptides may be immunogenic against, and/or derived from, any known arthropod borne disease. Examples of diseases, pathogens and vectors covered by the present invention, are set out in Table 1 above. Preferably, however, the relevant disease is malaria (including any malaria strain), as caused by any one of the strains of Plasmodium.

The specific sequences homologous to any of SEQ ID 1-134 described above are preferably the ones at the appropriate positions within known arthropod proteins, which can be found at the public NCBI protein database, which may be accessed online at the following URL address http://www.ncbi.nlm.nih.gov/entrez/query/static/help/helpdoc.html#Protein. The list is typically in the form |version number (gi number)|database identification (e.g. gb for GenBank)|NCBI accession number|optional further information (e.g. the accession number of the nucleotide sequence from which the protein sequence is derived). The sequences The protein database contains sequence data from the translated coding regions from DNA sequences in GenBank, EMBL, and DDBJ as well as protein sequences submitted to Protein Information Resource (PIR), SWISS-PROT, Protein Research Foundation (PRF), and Protein Data Bank (PDB) (sequences from solved structures).

The epitopes within the sequences defined above are not especially limited, provided that they contain 7 amino acid residues or more. Preferably the epitopes are at least of a length that is appropriate for the smaller immunogenic epitopes, such as CTL, T helper and B cell epitopes in a particular vertebrate species, such as in a human,. Typically the epitopes contain 8, 9, 10, or 11 amino acid residues, but may contain more if desired.

Although it may comprise more amino acids typically, the polypeptide comprises 100 amino acids or less, preferably between 7 and 100 amino acids, and more preferably from 8-75 amino acids. The size should not be so great that useful epitopes suffer from competition with non-protective epitopes in the immune system (for this reason full proteins are not included), nor should the size be so small that only a very narrow range of protection is offered. More preferred ranges are from 15-75 amino acids, 20-55 amino acids and 23-50 amino acids. It is particularly preferred that the polypeptide consists of (or substantially consists of) a sequence selected from the sequences defined above.

In addition to the polypeptides described above, the invention also provides multi-epitope immunogenic polypeptides comprising two or more polypeptides of the present invention either as multi-branched polypeptides or concatenated sequences. These multi-epitope polypeptides are not limited in size and may comprise e.g. up to 1400, or up to 900, or up to 550 amino acids. Thus, they extend not only to the polypeptides outlined above, but also to larger polypeptides, provided that these larger polypeptides comprise two or more units, each unit consisting of a polypeptide of the invention. Thus, a polypeptide having 100 repeating units of a 7-mer according to the present invention is encompassed by the present invention, as is a polypeptide having, say 52 units of one 8-mer epitope, and 23 units of a second 10-mer epitope. Polypeptides of this type will not suffer from the competition problems associated with similar length polypeptides that comprise only one or two epitopes. For the avoidance of doubt, the multi-epitope polypeptide may comprise multiple copies of the same epitope, or single copies of a plurality of different epitopes, or multiple copies of 2 or more epitopes. It is particularly preferred that a multi-epitope polypeptide comprises two or more of the sequences described above in SEQ ID 1-44 (and especially those in SEQ ID 1-6, 7, 8, 11, 12, 15, 16, 20, 26, 28, 30-32, 35, 41, and 42) or in SEQ ID 86-134.

As has been mentioned, the invention provides a polypeptide composition comprising one or more, preferably two or more different polypeptides as defined above. Thus, the polypeptide composition may comprise any number of polypeptides of the present invention together in the same sequence, mixture or formulation. The presence of a plurality of polypeptides together is useful since each may elicit its own immune response, widening the protective effect of the composition. It is particularly preferred that the composition contains two or more of (or all of) the sequences of SEQ ID 1-44 (and especially those in SEQ ID 1-6, 7, 8, 11, 12, 15, 16, 20, 26, 28, 30-32, 35, 41, and 42), and/or two or more of the epitopes within SEQ ID 86-134. In the composition each sequence and/or epitope may be present either as a separate peptide, or as a number of larger peptides comprising several concatenated epitopes and/or sequences (e.g. three sequences concatenated in one larger peptide and another 4 in another larger peptide, etc.).

The invention also provides a polypeptide construct, which construct comprises a polypeptide as defined above and a carrier. The construct may be formed by combining one or more epitopes and/or polypeptides as defined above with the carrier. The carrier may be a molecule, such as an adjuvant and/or an excipient. Combining in this context means either mixing together, or attaching together (e.g. via a covalent linkage).

The present invention further provides a polypeptide as defined above for use in medicine. Also provided is a medicament or vaccine composition against arthropod borne diseases, comprising a polypeptide as defined above, and one or more appropriate excipients and/or adjuvants, or a polypeptide construct as defined above and optionally one or more appropriate excipients and/or adjuvants (if the carrier part of the construct is itself an excipient or adjuvant, then a further excipient or adjuvant may not be needed). The excipient or adjuvant is not especially limited, and any excipients or adjuvants used in medicaments and vaccines may be employed. The medicament or vaccine composition may be produced according to any known method appropriately adapted to the present invention, such as by mixing a polypeptide of the invention with an appropriate excipient.

A method of producing a polypeptide as defined above is also provided by the invention. The method is not especially limited, and typically comprises joining two or more epitopes to form the polypeptide. The polypeptide may, however, be synthesised by direct chemical synthesis (e.g. incorporating one amino acid at a time until the full polypeptide is formed) or by recombinant methods. Such general methods are well known to the skilled person and may be adapted to the present invention as desired. In some instances, the polypeptide of the present invention may comprise additional amino acid sequences at one or both termini to help in synthesis of the polypeptide. These additional sequences are preferably from 1-5 amino acids in length. Typically 2 amino acids are involved. Examples of such sequences are provided as SEQ ID 86-129, as described above.

The invention still further provides use of a polypeptide or composition as defined above, in the manufacture of a medicament or vaccine, effective in the treatment or prevention of an arthropod borne disease. Also provided is a method of treating or preventing an arthropod borne disease, which method comprises administering a polypeptide, a composition, a medicament or a vaccine as defined above to a vertebrate. The method of administration is not especially limited, and may comprise subcutaneous, intramuscular, intra-venous, intra-dermal, or intra-nasal administration, or may be administered orally (e.g. in the form of a pill or a liquid preparation), or may be in the form of a suppository, if desired. The form of such administration preparations is not especially limited, and known forms may be employed with appropriate modifications that will be apparent to the skilled person. The dosage is not especially limited and may range from 1 pg to 100 g, preferably 1 ng to 100 g of the polypeptide per individual, depending upon the size, weight and species of the individual involved.

The invention may be applied to any vertebrate, since the immune systems of vertebrates operate in a related manner. Typically, the vertebrate referred to in the present context is a mammal, bird, a reptile or a fish. It is especially preferred that the vertebrate is a human, a domestic animal (such as a dog or a cat), a farm animal (such as a pig or a horse), a bovine animal (such as cattle), or fowl (such as a domestic bird, a farm bird, or a game bird). When the vertebrate is a bird, it is preferably a chicken, a turkey, a duck, or a goose.

Examples of human MHCs (HLAs) that may be associated with a particular T cell epitope in the present invention include the following:

HLA-A

A*010101, A*010102, A*010103, A*0102, A*0103, A*0104N, A*0106, A*0107, A*0108, A*0109, A*0110, A*02010101, A*02010102L, A*020102, A*020103, A*020104, A*020105, A*020106, A*020107, A*020108, A*020109, A*020110, A*020111, A*0202, A*020301, A*020302, A*0204, A*0205, A*020601, A*020602, A*020603, A*0207, A*0208, A*0209, A*0210, A*0211, A*0212, A*0213, A*0214, A*0215N, A*0216, A*021701, A*021702, A*0218, A*0219, A*022001, A*022002, A*0221, A*0222, A*0224, A*0225, A*0226, A*0227, A*0228, A*0229, A*0230, A*0231, A*0232N, A*0233, A*0234, A*023501, A*023502, A*0236, A*0237, A*0238, A*0239, A*0240, A*0241, A*0242, A*0243N, A*0244, A*0245, A*0246, A*0247, A*0248, A*0249, A*0250, A*0251, A*0252, A*0253N, A*0254, A*0255, A*0256, A*0257, A*0258, A*0259, A*0260, A*0261, A*0262, A*0263, A*0264, A*0265, A*0266, A*0267, A*0268, A*0269, A*0270, A*0271, A*0272, A*0273, A*03010101, A*03010102N, A*03010103, A*030102, A*030103, A*0302, A*0303N, A*0304, A*0305, A*0306, A*0307, A*0308, A*0309, A*0310, A*0311N, A*0312, A*0313, A*0314, A*110101, A*110102, A*1102, A*1103, A*1104, A*1105, A*1106, A*1107, A*1108, A*1109, A*1110, A*1111, A*1112, A*1113, A*1114, A*1115, A*1116, A*1117, A*1118, A*1119, A*2301, A*2302, A*2303, A*2304, A*2305, A*2306, A*2307N, A*2308N, A*2309, A*2310, A*2311N, A*2312, A*24020101, A*24020102L, A*240202, A*240203, A*240204, A*240205, A*240206, A*240301, A*240302, A*2404, A*2405, A*2406, A*2407, A*2408, A*2409N, A*2410, A*2411N, A*2413, A*2414, A*2415, A*2417, A*2418, A*2419, A*2420, A*2421, A*2422, A*2423, A*2424, A*2425, A*2426, A*2427, A*2428, A*2429, A*2430, A*2431, A*2432, A*2433, A*2434, A*2435, A*2436N, A*2437, A*2438, A*2439, A*2440N, A*2441, A*2442, A*2443, A*2444, A*2445N, A*2446, A*250101, A*250102, A*2502, A*2503, A*2504, A*2601, A*2602, A*2603, A*2604, A*2605, A*2606, A*260701, A*260702, A*2608, A*2609, A*2610, A*2611N, A*2612, A*2613, A*2614, A*2615, A*2616, A*2617, A*2618, A*2619, A*2620, A*2621, A*2622, A*2623, A*29010101, A*29010102N, A*290201, A*290202, A*290203, A*2903, A*2904, A*2905, A*2906, A*2907, A*2908N, A*2909, A*2910, A*2911, A*300101, A*300102, A*300201, A*300202, A*3003, A*3004, A*3006, A*3007, A*3008, A*3009, A*3010, A*3011, A*3012, A*310102, A*3102, A*3103, A*3104, A*3105, A*3106, A*3107, A*3108, A*3109, A*3110, A*3201, A*3202, A*3203, A*3204, A*3205, A*3206, A*3207, A*3208, A*3301, A*330301, A*330302, A*3304, A*3305, A*3306, A*3307, A*3401, A*3402, A*3403, A*3404, A*3405, A*3406, A*3601, A*3602, A*3603, A*3604, A*4301, A*6601, A*6602, A*6603, A*6604, A*680101, A*680102, A*680103, A*6802, A*680301, A*680302, A*6804, A*6805, A*6806, A*6807, A*6808, A*6809, A*6810, A*6811N, A*6812, A*6813, A*6814, A*6815, A*6816, A*6817, A*6818N, A*6819, A*6820, A*6821, A*6822, A*6823, A*6824, A*6825, A*6826, A*6827, A*6901, A*7401, A*7402, A*7403, A*7404, A*7405, A*7406, A*7407, A*7408, A*7409, A*7410, A*8001.

HLA-B

B*070201, B*070202, B*070203, B*070204, B*0703, B*0704, B*0705, B*0706, B*0707, B*0708, B*0709, B*0710, B*0711, B*0712, B*0713, B*0714, B*0715, B*0716, B*0717, B*0718, B*0719, B*0720, B*0721, B*0722, B*0723, B*0724, B*0725, B*0726, B*0727, B*0728, B*0729, B*0730, B*0731, B*0732, B*0733, B*0734, B*0735, B*0736, B*0737, B*0738, B*0801, B*0802, B*0803, B*0804, B*0805, B*0806, B*0807, B*0808N, B*0809, B*0810, B*0811, B*0812, B*0813, B*0814, B*0815, B*0816, B*0817, B*0818, B*0819N, B*0820, B*0821, B*0822, B*1301, B*1302, B*1303, B*1304, B*1306, B*1307N, B*1308, B*1309, B*1310, B*1311, B*1312, B*1313, B*1401, B*1402, B*1403, B*1404, B*1405, B*140601, B*140602, B*15010101, B*15010102N, B*150102, B*150103, B*150104, B*150105, B*1502, B*1503, B*1504, B*1505, B*1506, B*1507, B*1508, B*1509, B*1510, B*151101, B*151102, B*1512, B*1513, B*1514, B*1515, B*1516, B*15170101, B*15170102, B*1518, B*1519, B*1520, B*1521, B*1523, B*1524, B*1525, B*1526N, B*1527, B*1528, B*1529, B*1530, B*1531, B*1532, B*1533, B*1534, B*1535, B*1536, B*1537, B*1538, B*1539, B*1540, B*1542, B*1543, B*1544, B*1545, B*1546, B*1547, B*1548, B*1549, B*1550, B*1551, B*1552, B*1553, B*1554, B*1555, B*1556, B*1557, B*1558, B*1560, B*1561, B*1562, B*1563, B*1564, B*1565, B*1566, B*1567, B*1568, B*1569, B*1570, B*1571, B*1572, B*1573, B*1574, B*1575, B*1576, B*1577, B*1578, B*1579N, B*1580, B*1581, B*1582, B*1583, B*1584, B*1585, B*1586, B*1587, B*1588, B*1589, B*1590, B*1591, B*1592, B*1593, B*1594N, B*180101, B*180102, B*1802, B*1803, B*1804, B*1805, B*1806, B*1807, B*1808, B*1809, B*1810, B*1811, B*1812, B*1813, B*1814, B*1815, B*1817N, B*1818, B*1819, B*1820, B*2701, B*2702, B*2703, B*2704, B*270502, B*270503, B*270504, B*270505, B*270506, B*270507, B*2706, B*2707, B*2708, B*2709, B*2710, B*2711, B*2712, B*2713, B*2714, B*2715, B*2716, B*2717, B*2718, B*2719, B*2720, B*2721, B*2723, B*2724, B*2725, B*2726, B*350101, B*350102, B*3502, B*3503, B*3504, B*3505, B*3506, B*3507, B*3508, B*350901, B*350902, B*3510, B*3511, B*3512, B*3513, B*351401, B*351402, B*3515, B*3516, B*3517, B*3518, B*3519, B*3520, B*3521, B*3522, B*3523, B*3524, B*3525, B*3526, B*3527, B*3528, B*3529, B*3530, B*3531, B*3532, B*3533, B*3534, B*3535, B*3536, B*3537, B*3538, B*3539, B*3540N, B*3541, B*3542, B*3543, B*3544, B*3545, B*3546, B*3547, B*3548, B*3549, B*3550, B*3551, B*3552, B*3553N, B*3701, B*3702, B*3703N, B*3704, B*3705, B*3706, B*3707, B*3801, B*380201, B*380202, B*3803, B*3804, B*3805, B*3806, B*3807, B*3808, B*3809, B*3810, B*390101, B*390103, B*390104, B*390201, B*390202, B*3903, B*3904, B*3905, B*390601, B*390602, B*3907, B*3908, B*3909, B*3910, B*3911, B*3912, B*3913, B*3914, B*3915, B*3916, B*3917, B*3918, B*3919, B*3920, B*3922, B*3923, B*3924, B*3925N, B*3926, B*3927, B*3928, B*3929, B*3930, B*3931, B*3932, B*400101, B*400102, B*400103, B*400104, B*400105, B*400201, B*400202, B*4003, B*4004, B*4005, B*40060101, B*40060102, B*4007, B*4008, B*4009, B*4010, B*4011, B*4012, B*4013, B*401401, B*401402, B*401403, B*4015, B*4016, B*4018, B*4019, B*4020, B*4021, B*4022N, B*4023, B*4024, B*4025, B*4026, B*4027, B*4028, B*4029, B*4030, B*4031, B*4032, B*4033, B*4034, B*4035, B*4036, B*4037, B*4038, B*4039, B*4040, B*4042, B*4043, B*4044, B*4045, B*4046, B*4047, B*4048, B*4049, B*4050, B*4051, B*4052, B*4053, B*4054, B*4055, B*4056, B*4057, B*4101, B*4102, B*4103, B*4104, B*4105, B*4106, B*4201, B*4202, B*4204, B*420501, B*420502, B*4206, B*44020101, B*44020102S, B*440202, B*440203, B*440301, B*440302, B*4404, B*4405, B*4406, B*4407, B*4408, B*4409, B*4410, B*4411, B*4412, B*4413, B*4414, B*4415, B*4416, B*4417, B*4418, B*4419N, B*4420, B*4421, B*4422, B*4423N, B*4424, B*4425, B*4426, B*4427, B*4428, B*4429, B*4430, B*4431, B*4432, B*4433, B*4434, B*4435, B*4436, B*4437, B*4438, B*4439, B*4440, B*4501, B*4502, B*4503, B*4504, B*4505, B*4506, B*4507, B*4601, B*4602, B*4603, B*4604, B*47010101, B*47010102, B*4702, B*4703, B*4704, B*4705, B*4801, B*4802, B*4803, B*4804, B*4805, B*4806, B*4807, B*4808, B*4809, B*4810, B*4901, B*4902, B*4903, B*5001, B*5002, B*5004, B*510101, B*510102, B*510103, B*510104, B*510105, B*510201, B*510202, B*5103, B*5104, B*5105, B*5106, B*5107, B*5108, B*5109, B*5110, B*5111N, B*5112, B*511301, B*511302, B*5114, B*5115, B*5116, B*5117, B*5118, B*5119, B*5120, B*5121, B*5122, B*5123, B*5124, B*5126, B*5127N, B*5128, B*5129, B*5130, B*5131, B*5132, B*5133, B*5134, B*5135, B*5136, B*520101, B*520102, B*520103, B*520104, B*5202, B*5203, B*5204, B*5205, B*5206, B*530101, B*530102, B*5302, B*5303, B*5304, B*5305, B*5306, B*5307, B*5308, B*5309, B*5401, B*5402, B*5501, B*5502, B*5503, B*5504, B*5505, B*5507, B*5508, B*5509, B*5510, B*5511, B*5512, B*5513, B*5514, B*5515, B*5516, B*5601, B*5602, B*5603, B*5604, B*560501, B*560502, B*5606, B*5607, B*5608, B*5609, B*5610, B*5611, B*5612, B*5613, B*5614, B*570101, B*570102, B*5702, B*570301, B*570302, B*5704, B*5705, B*5706, B*5707, B*5708, B*5709, B*5801, B*5802, B*5804, B*5805, B*5806, B*5807, B*5808, B*5809, B*5810N, B*5901, B*670101, B*670102, B*6702, B*7301, B*7801, B*780201, B*780202, B*7803, B*7804, B*7805, B*8101, B*8102, B*8201, B*8202, B*8301.

HLA-C

Cw*010201, Cw*010202, Cw*0103, Cw*0104, Cw*0105, Cw*0106, Cw*0107, Cw*0108, Cw*0109, Cw*0110, Cw*020201, Cw*020202, Cw*020203, Cw*020204, Cw*020205, Cw*0203, Cw*0204, Cw*0205, Cw*0206, Cw*0207, Cw*0208, Cw*0209, Cw*030201, Cw*030202, Cw*030301, Cw*030302, Cw*030303, Cw*030304, Cw*030401, Cw*030402, Cw*030403, Cw*0305, Cw*0306, Cw*0307, Cw*0308, Cw*0309, Cw*0310, Cw*0311, Cw*0312, Cw*0313, Cw*0314, Cw*0315, Cw*0316, Cw*0317, Cw*0318, Cw*04010101, Cw*04010102, Cw*040102, Cw*0403, Cw*040401, Cw*040402, Cw*0405, Cw*0406, Cw*0407, Cw*0408, Cw*0409N, Cw*0410, Cw*0411, Cw*0412, Cw*0413, Cw*0414, Cw*0415, Cw*050101, Cw*050102, Cw*0502, Cw*0503, Cw*0504, Cw*0505, Cw*0506, Cw*0507N, Cw*0508, Cw*0509, Cw*0510, Cw*0602, Cw*0603, Cw*0604, Cw*0605, Cw*0606, Cw*0607, Cw*0608, Cw*0609, Cw*0610, Cw*0611, Cw*070101, Cw*070102, Cw*070103, Cw*07020101, Cw*07020102, Cw*07020103, Cw*0703, Cw*070401, Cw*070402, Cw*0705, Cw*0706, Cw*0707, Cw*0708, Cw*0709, Cw*0710, Cw*0711, Cw*0712, Cw*0713, Cw*0714, Cw*0715, Cw*0716, Cw*0717, Cw*0718, Cw*0719, Cw*0720, Cw*0721, Cw*0722, Cw*0723, Cw*0724, Cw*0725, Cw*0726, Cw*0727, Cw*0728, Cw*0729, Cw*080101, Cw*080102, Cw*0802, Cw*0803, Cw*0804, Cw*0805, Cw*0806, Cw*0807, Cw*0808, Cw*0809, Cw*0810, Cw*0811, Cw*0812, Cw*120201, Cw*120202, Cw*120203, Cw*120301, Cw*120302, Cw*120303, Cw*120401, Cw*120402, Cw*1205, Cw*1206, Cw*1207, Cw*1208, Cw*1209, Cw*1210, Cw*1211, Cw*1212, Cw*1213, Cw*1214, Cw*1215, Cw*140201, Cw*140202, Cw*140203, Cw*1403, Cw*1404, Cw*1405, Cw*150201, Cw*150202, Cw*1503, Cw*1504, Cw*150501, Cw*150502, Cw*150503, Cw*150504, Cw*1506, Cw*1507, Cw*1508, Cw*1509, Cw*1510, Cw*1511, Cw*1512, Cw*1601, Cw*1602, Cw*160401, Cw*1606, Cw*1701, Cw*1702, Cw*1703, Cw*1801, Cw*1802.

HLA-E

E*0101, E*010301, E*010302, E*010303, E*0104.

HLA-F

F*010101, F*010102.

HLA-G

G*010101, G*010102, G*010103, G*010104, G*010105, G*010106, G*010107, G*010108, G*0102, G*0103, G*010401, G*010402, G*010403, G*0105N, G*0106.

HLA-DRA

DRA*0101, DRA*010201, DRA*010202.

HLA-DRB1

DRB1*010101, DRB1*010102, DRB1*010103, DRB1*010201, DRB1*010202, DRB1*010203, DRB1*010204, DRB1*0103, DRB1*0104, DRB1*0105, DRB1*0106, DRB1*0107, DRB1*0108, DRB1*0109, DRB1*0110, DRB1*0111, DRB1*030101, DRB1*030102, DRB1*030201, DRB1*030202, DRB1*0303, DRB1*0304, DRB1*030501, DRB1*030502, DRB1*0306, DRB1*0307, DRB1*0308, DRB1*0309, DRB1*0310, DRB1*0311, DRB1*0312, DRB1*0313, DRB1*0314, DRB1*0315, DRB1*0316, DRB1*0317, DRB1*0318, DRB1*0319, DRB1*0320, DRB1*0321, DRB1*0322, DRB1*0323, DRB1*0324, DRB1*0325, DRB1*0326, DRB1*0327, DRB1*0328, DRB1*040101, DRB1*040102, DRB1*0402, DRB1*040301, DRB1*040302, DRB1*0404, DRB1*040501, DRB1*040502, DRB1*040503, DRB1*040504, DRB1*0406, DRB1*040701, DRB1*040702, DRB1*040703, DRB1*0408, DRB1*0409, DRB1*0410, DRB1*0411, DRB1*0412, DRB1*0413, DRB1*0414, DRB1*0415, DRB1*0416,

DRB1*0417, DRB1*0418, DRB1*0419, DRB1*0420, DRB1*0421, DRB1*0422, DRB1*0423, DRB1*0424, DRB1*0425, DRB1*0426, DRB1*0427, DRB1*0428, DRB1*0429, DRB1*0430, DRB1*0431, DRB1*0432, DRB1*0433, DRB1*0434, DRB1*0435, DRB1*0436, DRB1*0437, DRB1*0438, DRB1*0439, DRB1*0440, DRB1*0441, DRB1*0442, DRB1*0443, DRB1*0444, DRB1*0445, DRB1*0446, DRB1*0447, DRB1*0448, DRB1*0449, DRB1*0450, DRB1*070101, DRB1*070102, DRB1*0703, DRB1*0704, DRB1*0705, DRB1*0706, DRB1*0707, DRB1*0708, DRB1*080101, DRB1*080102, DRB1*080201, DRB1*080202, DRB1*080203, DRB1*080302, DRB1*080401, DRB1*080402, DRB1*080403, DRB1*080404, DRB1*0805, DRB1*0806, DRB1*0807, DRB1*0808, DRB1*0809, DRB1*0810, DRB1*0811, DRB1*0812, DRB1*0813, DRB1*0814, DRB1*0815, DRB1*0816, DRB1*0817, DRB1*0818, DRB1*0819, DRB1*0820, DRB1*0821, DRB1*0822, DRB1*0823, DRB1*0824, DRB1*0825, DRB1*0826, DRB1*0827, DRB1*0828, DRB1*0829, DRB1*090102, DRB1*090103, DRB1*0902, DRB1*0903, DRB1*100101, DRB1*100102, DRB1*110101, DRB1*110102, DRB1*110103, DRB1*110104, DRB1*110105, DRB1*1102, DRB1*1103, DRB1*110401, DRB1*110402, DRB1*1105, DRB1*110601, DRB1*110602, DRB1*1107, DRB1*110801, DRB1*110802, DRB1*1109, DRB1*1110, DRB1*1111, DRB1*111201, DRB1*111202, DRB1*1113, DRB1*1114, DRB1*1115, DRB1*1116, DRB1*1117, DRB1*1118, DRB1*1119, DRB1*1120, DRB1*1121, DRB1*1122, DRB1*1123, DRB1*1124, DRB1*1125, DRB1*1126, DRB1*112701, DRB1*112702, DRB1*1128, DRB1*1129, DRB1*1130, DRB1*1131, DRB1*1132, DRB1*1133, DRB1*1134, DRB1*1135, DRB1*1136, DRB1*1137, DRB1*1138, DRB1*1139, DRB1*1140, DRB1*1141, DRB1*1142, DRB1*1143, DRB1*1144, DRB1*1145, DRB1*1146, DRB1*1147, DRB1*1148, DRB1*1149, DRB1*1150, DRB1*1151, DRB1*1152, DRB1*1153, DRB1*1154, DRB1*120101, DRB1*120102, DRB1*120201, DRB1*120202, DRB1*120302, DRB1*1204, DRB1*1205, DRB1*1206, DRB1*1207, DRB1*1208, DRB1*1209, DRB1*1210, DRB1*130101, DRB1*130102, DRB1*130103, DRB1*130201, DRB1*130202, DRB1*130301, DRB1*130302, DRB1*1304, DRB1*1305, DRB1*1306, DRB1*130701, DRB1*130702, DRB1*1308, DRB1*1309, DRB1*1310, DRB1*1311, DRB1*1312, DRB1*1313, DRB1*131401, DRB1*131402, DRB1*1315, DRB1*1316, DRB1*1317, DRB1*1318, DRB1*1319, DRB1*1320, DRB1*1321, DRB1*1322, DRB1*1323, DRB1*1324,

DRB1*1325, DRB1*1326, DRB1*1327, DRB1*1328, DRB1*1329, DRB1*1330, DRB1*1331, DRB1*1332, DRB1*1333, DRB1*1334, DRB1*1335, DRB1*1336, DRB1*1337, DRB1*1338, DRB1*1339, DRB1*1340, DRB1*1341, DRB1*1342, DRB1*1343, DRB1*1344, DRB1*1345, DRB1*1346, DRB1*1347, DRB1*1348, DRB1*1349, DRB1*1350, DRB1*1351, DRB1*1352, DRB1*1353, DRB1*1354, DRB1*1355, DRB1*1356, DRB1*1357, DRB1*1358, DRB1*1359, DRB1*1360, DRB1*1361, DRB1*1362, DRB1*1363, DRB1*1364, DRB1*1365, DRB1*140101, DRB1*140102, DRB1*1402, DRB1*140301, DRB1*140302, DRB1*1404, DRB1*140501, DRB1*140502, DRB1*1406, DRB1*140701, DRB1*140702, DRB1*1408, DRB1*1409, DRB1*1410, DRB1*1411, DRB1*1412, DRB1*1413, DRB1*1414, DRB1*1415, DRB1*1416, DRB1*1417, DRB1*1418, DRB1*1419, DRB1*1420, DRB1*1421, DRB1*1422, DRB1*1423, DRB1*1424, DRB1*1425, DRB1*1426, DRB1*1427, DRB1*1428, DRB1*1429, DRB1*1430, DRB1*1431, DRB1*1432, DRB1*1433, DRB1*1434, DRB1*1435, DRB1*1436, DRB1*1437, DRB1*1438, DRB1*1439, DRB1*1440, DRB1*1441, DRB1*1442, DRB1*1443, DRB1*1444, DRB1*1445, DRB1*1446, DRB1*1447, DRB1*1448, DRB1*150101, DRB1*150102, DRB1*150103, DRB1*150104, DRB1*150105, DRB1*150201, DRB1*150202, DRB1*150203, DRB1*1503, DRB1*1504, DRB1*1505, DRB1*1506, DRB1*1507, DRB1*1508, DRB1*1509, DRB1*1510, DRB1*1511, DRB1*1512, DRB1*1513, DRB1*1514, DRB1*1515, DRB1*1516, DRB1*160101, DRB1*160102, DRB1*160201, DRB1*160202, DRB1*1603, DRB1*1604, DRB1*160501, DRB1*160502, DRB1*1607, DRB1*1608.

HLA-DRB2-9

DRB2*0101, DRB3*010101, DRB3*01010201, DRB3*01010202, DRB3*010103, DRB3*010104, DRB3*0102, DRB3*0103, DRB3*0104, DRB3*0105, DRB3*0106, DRB3*0107, DRB3*0108, DRB3*0109, DRB3*0110, DRB3*0111, DRB3*0201, DRB3*020201, DRB3*020202, DRB3*020203, DRB3*020204, DRB3*0203, DRB3*0204, DRB3*0205, DRB3*0206, DRB3*0207, DRB3*0208, DRB3*0209, DRB3*0210, DRB3*0211, DRB3*0212, DRB3*0213, DRB3*0214, DRB3*0215, DRB3*0216, DRB3*0217, DRB3*0218, DRB3*0219, DRB3*030101, DRB3*030102, DRB3*0302, DRB3*0303, DRB4*01010101, DRB4*0102, DRB4*01030101, DRB4*01030102N, DRB4*010302, DRB4*010303, DRB4*010304, DRB4*0104, DRB4*0105, DRB4*0106,

DRB4*0107, DRB4*0201N, DRB4*0301N, DRB5*010101, DRB5*010102, DRB5*0102, DRB5*0103, DRB5*0104, DRB5*0105, DRB5*0106, DRB5*0107, DRB5*0108N, DRB5*0109, DRB5*0110N, DRB5*0111, DRB5*0112, DRB5*0113, DRB5*0202, DRB5*0203, DRB5*0204, DRB5*0205, DRB6*0101, DRB6*0201, DRB6*0202, DRB7*010101, DRB7*010102, DRB8*0101, DRB9*0101.

HLA-DQA1

DQA1*010101, DQA1*010102, DQA1*010201, DQA1*010202, DQA1*0103, DQA1*010401, DQA1*010402, DQA1*0105, DQA1*0106, DQA1*0107, DQA1*0201, DQA1*030101, DQA1*0302, DQA1*0303, DQA1*040101, DQA1*040102, DQA1*0402, DQA1*0403N, DQA1*0404, DQA1*050101, DQA1*050102, DQA1*0502, DQA1*0503, DQA1*0504, DQA1*0505, DQA1*060101, DQA1*060102, DQA1*0602.

HLA-DQB1

DQB1*020101, DQB1*020102, DQB1*0202, DQB1*0203, DQB1*030101, DQB1*030102, DQB1*030201, DQB1*030202, DQB1*030302, DQB1*030303, DQB1*0304, DQB1*030501, DQB1*030502, DQB1*030503, DQB1*0306, DQB1*0307, DQB1*0308, DQB1*0309, DQB1*0310, DQB1*0311, DQB1*0312, DQB1*0313, DQB1*0401, DQB1*0402, DQB1*050101, DQB1*050102, DQB1*050201, DQB1*050202, DQB1*050301, DQB1*050302, DQB1*0504, DQB1*060101, DQB1*060102, DQB1*060103, DQB1*0602, DQB1*0603, DQB1*060401, DQB1*060402, DQB1*060501, DQB1*060502, DQB1*0606, DQB1*0607, DQB1*0608, DQB1*0609, DQB1*0610, DQB1*061101, DQB1*061102, DQB1*0612, DQB1*0613, DQB1*0614, DQB1*0615, DQB1*0616, DQB1*0617, DQB1*0618, DQB1*0619, DQB1*0620, DQB1*0621, DQB1*0622, DQB1*0623.

HLA-DPA1

DPA1*010301, DPA1*010302, DPA1*010303, DPA1*0104, DPA1*0105, DPA1*0106, DPA1*0107, DPA1*0108, DPA1*020101, DPA1*020102, DPA1*020103, DPA1*020104, DPA1*020105, DPA1*020106, DPA1*020201, DPA1*020202, DPA1*020203, DPA1*0203, DPA1*0301, DPA1*0302, DPA1*0303, DPA1*0401.

HLA-DPB1

DPB1*010101, DPB1*010102, DPB1*010103, DPB1*0102, DPB1*020102, DPB1*020103, DPB1*020104, DPB1*020105, DPB1*020106, DPB1*0202, DPB1*0203, DPB1*030101, DPB1*030102, DPB1*0302, DPB1*040101, DPB1*040102, DPB1*0402, DPB1*0501, DPB1*0601, DPB1*0801, DPB1*0901, DPB1*1001, DPB1*110101, DPB1*110102, DPB1*1301, DPB1*1401, DPB1*1501, DPB1*1601, DPB1*1701, DPB1*1801, DPB1*1901, DPB1*200101, DPB1*200102, DPB1*2101, DPB1*2201, DPB1*2301, DPB1*2401, DPB1*2501, DPB1*260101, DPB1*260102, DPB1*2701, DPB1*2801, DPB1*2901, DPB1*3001, DPB1*3101, DPB1*3201, DPB1*3301, DPB1*3401, DPB1*3501, DPB1*3601, DPB1*3701, DPB1*3801, DPB1*3901, DPB1*4001, DPB1*4101, DPB1*4401, DPB1*4501, DPB1*4601, DPB1*4701, DPB1*4801, DPB1*4901, DPB1*5001, DPB1*5101, DPB1*5201, DPB1*5301, DPB1*5401, DPB1*5501, DPB1*5601, DPB1*5701, DPB1*5801, DPB1*5901, DPB1*6001, DPB1*6101N, DPB1*6201, DPB1*6301, DPB1*6401N, DPB1*6501, DPB1*6601, DPB1*6701, DPB1*6801, DPB1*6901, DPB1*7001, DPB1*7101, DPB1*7201, DPB1*7301, DPB1*7401, DPB1*7501, DPB1*7601, DPB1*7701, DPB1*7801, DPB1*7901, DPB1*8001, DPB1*8101, DPB1*8201, DPB1*8301, DPB1*8401, DPB1*8501, DPB1*8601, DPB1*8701, DPB1*8801, DPB1*8901, DPB1*9001, DPB1*9101, DPB1*9201, DPB1*9301, DPB1*9401, DPB1*9501, DPB1*9601, DPB1*9701, DPB1*9801, DPB1*9901.

HLA-DMA

DMA*0101, DMA*0102, DMA*0103, DMA*0104.

HLA-DMB

DMB*0101, DMB*0102, DMB*0103, DMB*0104, DMB*0105, DMB*0106.

HLA-DOA

DOA*010101, DOA*01010201, DOA*01010202, DOA*01010203, DOA*010103, DOA*01010401, DOA*01010402, DOA*010105.

HLA-DOB

DOB*01010101, DOB*01010102, DOB*010102, DOB*010201, DOB*010202, DOB*0103, DOB*01040101, DOB*01040102.

MHC Class I

H-2Db, H-2Dd, H-2Dk, H-2Dq, H-2Kb, H-2Kd, H-2Kk, H-2Ld, H-2M3, H-2Ad, H-2Ag7, H-2Ak, H2-Ab, H-2Ed, H-2Ek, H-2Bxk, H-2F, H-2I, H-2P, H-2R, H-2S, H-2Sxd, H-2T4, H-2U.

MHC Class II

I-Ab, I-Ad, I-Ag7, I-Ak, I-Ap, I-Aq, I-Ar, I-As, I-Au, I-Av, I-Ea, I-Eb, I-Ed, I-Ek, I-Es, I-Eu, H-2Q, H-2Qa-2, H-2Qa-2a, Qa-1a, Qa-1b.

The invention is not limited to such MHC and HLA molecules, and can be adapted to newly discovered such molecules, if desired, simply by establishing the reactivity of substances such as peptides with the molecules. This can be readily achieved using known techniques that are standard in the field. Particularly preferred HLA alleles for use with the present invention include the following:

HLA Class I

HLA A	HLA B	HLA Cw
A*6802	B*5801	Cw*1701
A*6801	B*5701	Cw*1601
A*6601	B*5501	Cw*1502
A*3303	B*5201	Cw*1402
A*3301	B*5101	Cw*1203
A*3201	B*5001	Cw*0802
A*310102	B*4901	Cw*0801
A*3002	B*4501	Cw*0704
A*3001	B*4403	Cw*0703
A*2902	B*4402	Cw*0702
A*2608	B*4101	Cw*0701
A*2601	B*4002	Cw*0602
A*2501	B*4001	Cw*0501
A*2402	B*3901	Cw*0401
A*2301	B*3801	Cw*0304
A*1101	B*3701	Cw*0303
A*0302	B*3503	Cw*0202
A*0301	B*3501	Cw*0102
A*0205	B*2705	
A*0201	B*1801	
A*0101	B*1501	
	B*1402	
	B*1401	
	B*1302	
	B*0801	
	B*0705	
	B*0702	

HLA Class II

HLA DPB	HLA DQA	HLA DQB	HLA DRB
DPB1*1701	DQA1*0505	DQB1*0604	DRB1*1601
DPB1*1301	DQA1*0501	DQB1*0603	DRB1*1501
DPB1*1001	DQA1*0401	DQB1*0602	DRB1*1401
DPB1*0601	DQA1*0303	DQB1*0503	DRB1*1302
DPB1*0501	DQA1*0302	DQB1*0502	DRB1*1301
DPB1*0402	DQA1*0301	DQB1*0501	DRB1*1201
DPB1*0401	DQA1*0201	DQB1*0402	DRB1*1104
DPB1*0301	DQA1*0104	DQB1*0303	DRB1*1101
DPB1*0201	DQA1*0103	DQB1*0302	DRB1*0801
DPB1*0101	DQA1*0102	DQB1*0301	DRB1*0701
	DQA1*0101	DQB1*0202	DRB1*0404
		DQB1*0201	DRB1*0401
			DRB1*0301
			DRB1*0103
			DRB1*0102
			DRB1*0101

The most preferred alleles according to the invention are the following:

HLA-A*0201, HLA-A*0206, HLA-A*0301, HLA-A*1101, HLA-A*2402, HLA-A*3401, HLA-B*0702, HLA-B*0801, HLA-B*1301, HLA-B*27, HLA-B*4002, HLA-B*5101, HLA-Cw*03, HLA-cW*07

HLA-DRB1*0301, HLA-DRB1*0401, HLA-DRB1*0701, HLA-DRB1*1501, HLA-DRB1*1104, HLA-DRB1*1101, HLA-DRB4*0101

HLA-DQA1*01, HLA-DQA1*02, HLA-DQA1*05

HLA-DQB1*03, HLA-DQB1*04, HLA-DQB1*05, HLA-DQB1*06

HLA-DPA1*01, HLA-DPA1*02

HLA-DPB1*02, HLA-DPB1*04

The invention will now be described by way of example only, with reference to the following specific embodiments.

EXAMPLES

Preparation of arthropod saliva protein fractions

In order to determine the effect of specific sequences of the invention, their immunogenicity may be tested against various arthropod saliva protein fractions. Those sequences that cause a vertebrate to produce immune system cells that recognise at least one epitope in specific saliva protein fractions are useful in the vaccines of the present invention.

The saliva protein fractions can easily be isolated using standard laboratory techniques, which are well known to the skilled person. Any arthropod saliva protein fractions may be used, since the inventors have determined that it is the mass of the protein fraction that is important. The fractions of mass 40kDa or less, 30kDa or less, preferably from 20-40kDa, and more preferably 20kDa or less, are particularly useful.

The following protocol is provided to exemplify the protein fraction against which candidate sequences may be tested. It utilises saliva from anopheles gambiae mosquitoes, although any arthropod saliva may be employed.

Anopheles gambiae salivary gland pairs (SGP) were dissected from female mosquitoes colonies. Fifteen SGPs were collected in 20 μl of PBS and lysed by adding 5 μl of 5x SDS-PAGE sample buffer containing 0.25% 2-β-ME. After vortexing and boiling for 5 min, the protein mixture was loaded on a Novex 4-20% gradient Tris-Glycine gel (INVITROGEN). The gel was then silver stained and photographed (see Figure 9).

The dashed arrows indicate the position of the cut off points for selection of SGP fractions that may be used for immunisation and study (i.e. <20 kDa, 20-40 kDa, 40-80 kDa and >80 kDa). The solid arrow indicates the position of the cut off point for selection of SGP fractions of <30 kDa and >30 kDa.

Identifying candidate protein samples from mosquito saliva

Taking a similar approach to the exemplary protocol outlined above, several saliva protein samples were prepared for study.

Salivary glands (SG) were dissected from female Anopheles gambiae mosquitoes and stored in PBS at -70°C until use.

For gel analysis, SGs were lysed by freeze-thawing, Novex® IEF Sample Buffer pH 3-10 (Invitrogen) added and the resulting material analysed in a Novex® IEF Gel (Invitrogen). The gel was then fixed with 12% TCA, washed three times in water and stained with Comassie Blue. The resulting gel is shown in Figure 1.

After staining and destaining, the IEF gel was incubated in 20% ethanol for 10 minutes and the gel strip containing the lane with the SG proteins cut out. This gel strip was equilibrated for 5 min in 2x SDS-PAGE sample buffer containing 20% ethanol, rinsed twice in SDS-PAGE sample buffer and loaded in the single well of a Novex® 4-20% Tris-Glycine Gel (Invitrogen). The resulting gel is shown in Figure 2.

The gel containing the separated SG proteins was finally stained using the Proteosilver stain kit (Sigma) according to the manufacturer's instructions.

From the gel, four groups of proteins (designated compounds 1, 2, 3 and 4) were identified for analysis:

Compound 1	*Salivary Gland Protein Fraction <20 kDa*
Compound 2	*Salivary Gland Protein Fraction 20 kDa < X < 40 kDa*
Compound 3	*Salivary Gland Protein Fraction 40 kDa < X < 80 kDa*
Compound 4	*Salivary Gland Protein Fraction >80 kDa*

EXPERIMENT 1

Aims:

(A) Establish efficacy of candidates in protecting animals against challenge by bite of infected mosquitoes (i.e. Plasmodium yoelii nigeriensis infected Anopheles gambiae)

(B) Establish cross-reactivity level of the anti-mosquito response induced by candidates across different species of mosquitoes (e.g. Anopheles gambiae and Anopheles stephensi).
(C) Establish efficacy of candidates in preventing infection of mosquitoes (Anopheles gambiae and Anopheles stephensi) by Plasmodium yoelii nigeriensis after biting immunised and infected mice.

Candidate compounds

The compounds selected were those identified above:

Compound 1	*Salivary Gland Protein Fraction <20 kDa*
Compound 2	*Salivary Gland Protein Fraction 20 kDa < X < 40 kDa*
Compound 3	*Salivary Gland Protein Fraction 40 kDa < X < 80 kDa*
Compound 4	*Salivary Gland Protein Fraction >80 kDa*

Strain and number of animals involved:

CD1 mice are used. There were five experimental groups (1, 2, 3, 4 and 5) with group 1 being the negative control group and groups 2-5 the test groups, each group having nine (9) animals.

Experimental protocol

Day 1: 4 groups were immunised (groups 2, 3, 4 and 5) of 9 CD1 mice each (N = 4 x 9 = 36) with subcutaneous doses of candidate vaccine compounds (Group 2 with compound 1, Group 3 with compound 2, and so on).

Day 14: All animals were boosted with the same doses of candidate vaccine candidate vaccine compounds (Group 2 with compound 1, Group 3 with compound 2, and so on).

Day 21: All animals were test bled. Samples were stored frozen (-20°C) until collection. Each group was split into further subgroups: A – 5 animals, B – 4 animals.

Group 1A – 5 animals	Group 1B – 4 animals
Group 2A – 5 animals	Group 2B – 4 animals
Group 3A – 5 animals	Group 3B – 4 animals

Group 4A – 5 animals Group 4B – 4 animals
Group 5A – 5 animals Group 5B – 4 animals

Subgroups A

Day 28: All animals in subgroups A were challenged via bite of 5-9 infected mosquitoes (i.e. *Plasmodium yoelii nigeriensis* infected *Anopheles gambiae*) in the belly area. All animals were maintained until parasitemia is first established or for a maximum of 6 weeks after challenge with infected mosquitoes. All animals were killed by exsanguination and the sera samples were stored frozen (-20°C) until collection.

Subgroups B

Day 28: All animals in subgroups B were used to feed (in the belly area) the following number of fresh (uninfected) mosquitoes:

5-10 *Anopheles gambiae* AND 5-10 *Anopheles stephensi*

All these mosquitoes were tested for:

1. Survival over an 8 day period.
2. Number of eggs laid.
3. Number of eggs produced.
4. Number of adults (F1) produced from the eggs laid.

Day 32: All animals in subgroups B were infected with *Plasmodium yoelii nigeriensis* by direct IV inoculation of parasites.

Day 32: Once active malaria infection had been identified in all (or at least 75%) animals in Group B, all infected animals were used to feed (in the belly area) large numbers (>10 per mice) of both fresh *Anopheles gambiae* AND fresh *Anopheles stephensi*.

All these mosquitoes were tested for:

1. Survival over an 8 day period.

2. Number of mosquitoes with malarial parasites in their salivary glands amongst those who survived the required incubation period (17 days).

3. Number of eggs laid.
4. Number of eggs produced.
5. Number of adults (F1) produced from the eggs laid.

After mosquito feeding, all animals were killed by exsanguination and the sera samples were stored frozen (-20°C) until collection.

Any volume of the experimental compound remaining at the end of the study was stored frozen (-20°C) until collection.

Results

Graphical representations of the results of the experiments are shown in Figures 3A-3I and 4. The first set of Figures (Figures 3x) show data on the effect of the vaccine on the fecundity of mosquitoes, for each of the Groups 1-5, as follows:

3A : Per cent that fed
3B : Mean no eggs produced
3C : Mean no eggs laid
3D : Per cent hatch rate
3E : Mean no larvae
3F : Mean no pupae
3G : Per cent pupation
3H : Per cent emergence
3I : Mean no adults

The numerical data underpinning each of these graphical representations A-I are set down respectively (column A for Figure 3A, and so on) in Table 2 below:

Table 2 – Data for Figures 3A-3I

Mean ± Std. Error	A	B	C	D	E
Group 1	82.7 ± 4	29.6 ± 5	37.5 ± 4.6	20.1 ± 3.9	54.6 ± 4.7
Group 2	66.7 ± 6.5	17.1 ± 3.8	12 ± 3	9.7 ± 2.4	51.9 ± 7.4
Group 3	81.2 ± 3.4	19.1 ± 4.6	16 ± 3.7	12.9 ± 4	44.3 ± 8.7
Group 4	73.7 ± 4.9	24.1 ± 3.9	26.4 ± 3.9	15 ± 2.8	56.6 ± 6.6
Group 5	75.7 ± 4.5	30.1 ± 4.1	32.4 ± 5	16.9 ± 2.5	55 ± 4.6

Mean ± Std. Error	F	G	H	I
Group 1	15.6 ± 3	82.9 ± 3	14 ± 2.8	92.4 ± 2.6
Group 2	7.8 ± 1.8	86.1 ± 5.7	6.9 ± 1.8	80 ± 7.8
Group 3	12.6 ± 3.7	77.7 ± 7.3	12.5 ± 3.6	92 ± 1.8
Group 4	12.3 ± 2.1	88.5 ± 3.3	10.8 ± 1.9	88.8 ± 2.4
Group 5	14.4 ± 2.2	88.4 ± 2.7	12.1 ± 1.8	84.8 ± 3.7

The data set out in Tables 3A-3I below show the p-values obtained using a Mann-Whitney non-parametric statistical analysis of the above data.

Table 3A – P values for per cent that fed

	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	0.10504			
Group 3	0.41727	0.14813		
Group 4	0.14813	0.26543	0.10504	
Group 5	0.23235	0.23235	0.26543	0.5

Table 3B - P values for mean no eggs produced

	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	4.6E-05			
Group 3	0.00042	0.1612		
Group 4	0.04038	0.0032	0.01602	
Group 5	0.17836	0.0018	0.00551	0.22172

Table 3C - P values for mean no eggs laid

	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	0.09451			
Group 3	0.07626	0.49308		
Group 4	0.27329	0.09416	0.10891	
Group 5	0.30616	0.01171	0.00879	0.15145

Table 3D - P values for per cent hatch rate

	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	0.46767			
Group 3	0.19155	0.29445		
Group 4	0.36699	0.33803	0.14632	
Group 5	0.40058	0.47907	0.15041	0.40187

Table 3E - P values for mean no larvae

	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	0.04897			
Group 3	0.07371	0.45038		
Group 4	0.29801	0.06197	0.11695	
Group 5	0.4499	0.0164	0.06524	0.23691

Table 3F - P values for mean no pupae

	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	0.06224			
Group 3	0.28662	0.22088		
Group 4	0.44057	0.04081	0.35186	
Group 5	0.40546	0.01118	0.17373	0.25066

Table 3G - P values for per cent pupation

	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	0.11277			
Group 3	0.3664	0.13511		
Group 4	0.11848	0.50798	0.07764	
Group 5	0.10627	0.41971	0.08354	0.37388

Table 3H - P values for per cent emergence

	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	0.06876			
Group 3	0.12815	0.25697		
Group 4	0.059	0.33712	0.2767	
Group 5	0.02006	0.4853	0.18532	0.2943

Table 3I - P values for mean no adults

	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	0.04391			
Group 3	0.47127	0.06735		
Group 4	0.3981	0.03908	0.47468	
Group 5	0.46488	0.0123	0.31328	0.24603

Figure 4 shows the per cent survival of the mosquitoes in each of the five groups, over a period of 8 days. The number of mosquitoes in each group at day zero was as follows:

Group 1 (control):	100
Group 2:	89
Group 3:	100
Group 4:	99
Group 5:	98

The data relevant to the graphical representation of Figure 4 is set out in Table 4 below:

Table 4 – Data for Figure 4

Mean ± Std. Error	Day 0	Day 1	Day 2	Day 3	Day 4
Group 1	100 ± 0	86 ± 4.2	77 ± 5	73 ± 5.2	61 ± 4.6
Group 2	100 ± 0	53.4 ± 5.5	40.4 ± 5.3	36.4 ± 4.9	28 ± 5.2
Group 3	100 ± 0	67 ± 4.5	51 ± 4.7	44 ± 4.9	36 ± 4.4
Group 4	100 ± 0	78 ± 5.1	62 ± 5.7	60 ± 5.6	51 ± 4.9
Group 5	100 ± 0	73 ± 5.6	53.4 ± 6.7	48.4 ± 6.3	41.4 ± 5.6

Mean ± Std. Error	Day 5	Day 6	Day 7	Day 8
Group 1	58 ± 4.8	56 ± 4.4	53 ± 4.5	53 ± 4.5
Group 2	23.7 ± 5	21.4 ± 4.8	20 ± 4.8	20 ± 4.8
Group 3	30 ± 4.1	28 ± 3.7	28 ± 3.7	28 ± 3.7
Group 4	46 ± 5	46 ± 5	43 ± 5.2	43 ± 5.2
Group 5	36.7 ± 5	33.7 ± 5.3	29.4 ± 5	29.4 ± 5

The data set out in Table 5(i)-5(viii) below show the p-values obtained using a Mann-Whitney non-parametric statistical analysis of the above data.

Table 5(i) – P values for day 1

Day 1	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	2.6E-05			
Group 3	0.00217	0.0326		
Group 4	0.13228	0.0007	0.0387	
Group 5	0.0526	0.0071	0.1588	0.2738

Table 5(ii) – P values for day 2

Day 2	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	1.1E-05			
Group 3	0.0003	0.0676		
Group 4	0.0274	0.00453	0.06759	
Group 5	0.0049	0.07833	0.44598	0.166

Table 5(iii) – P values for day 3

Day 3	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	1.1E-05			
Group 3	0.0002	0.20756		
Group 4	0.04955	0.00253	0.0213	
Group 5	0.00311	0.07692	0.26734	0.10538

Table 5(iv) – P values for day 4

Day 4	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	3.8E-05			
Group 3	0.00032	0.12615		
Group 4	0.09515	0.00175	0.01525	
Group 5	0.0065	0.04759	0.24854	0.09848

Table 5(v) – P values for day 5

Day 5	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	2.6E-05			
Group 3	6.9E-05	0.14515		
Group 4	0.07976	0.00198	0.00944	
Group 5	0.00393	0.03955	0.14962	0.09848

Table 5(vi) – P values for day 6

Day 6	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	1E-05			
Group 3	2.3E-05	0.12817		
Group 4	0.12022	0.00081	0.00393	
Group 5	0.0034	0.05802	0.22172	0.0558

Table 5(vii) – P values for day 7

Day 7	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	2.3E-05			
Group 3	8.4E-05	0.07147		
Group 4	0.13648	0.00175	0.01315	
Group 5	0.00132	0.09032	0.40795	0.03713

Table 5(viii) – P values for day 8

Day 8	Compound 1	Compound 2	Compound 3	Compound 4
Group 1				
Group 2	2.3E-05			
Group 3	8.4E-05	0.07147		
Group 4	0.13648	0.00175	0.01315	
Group 5	0.00132	0.09032	0.40795	0.03713

EXPERIMENT 2: Induction of cytokine production by polypeptide antigens immunisation

Peptides and recombinant proteins

From experiment 1, the most effective fraction was studied further, to identify polypeptides for that may be employed in vaccines. The utility of the peptides of the present invention was determined using the following protocol.

Immunisations

All the polypeptides under study (antigen preparations) are synthesised by Fmoc chemistry.

Six to ten week old C57BL/6 mice are immunised subcutaneously with a 200 µl dose of the antigen preparation per mouse. In the test group, each dose of the antigen preparation contains an equimolar mixture of the peptides (10 nmol each) prepared in adjuvant (Sigma) according to the manufacturer's instructions. In the control group, each dose of the antigen preparation contains an equivalent dose of a non-relevant polypeptide prepared in IFA (Sigma) according to the manufacturer's instructions (NRP preparation).

On day 15 post-immunisation, all animals receive a booster immunisation using the same doses and route of delivery as originally.

Finally, on day 20 all animals are culled and their spleens and sera are collected.

Cytokine ELISA

Mouse spleens belonging to the same experimental group are pooled, gently pressed through cell strainers and red blood cells removed by treatment with red cell lysis buffer (nine parts 0.16 M NH_4Cl and one part of 0.17 M Tris, pH 7.2). Splenocyte suspensions from each experimental group are plated in 96-well plates in quadruplicate at a density of 4×10^6 cells/well in IMDM medium (Invitrogen) supplemented with 0.02 mM β-mercaptoethanol (Sigma), 50 IU/50 mg/ml of penicillin/streptomycin (Sigma) and 10 % FCS (Sigma) and containing each of the polypeptide antigens under study (2 μM). After 3 days incubation at 37°C, the supernatant is collected and analysed for IFN-γ and IL-4 by a sandwich cytokine ELISA according to the manufacturer's protocol (Pharmingen). The lower detection limits for the assay are 9.77 pg/ml for IL-4 and 39.06 pg/ml for IFN-γ.

IgG2a specific ELISA

Microtiter ELISA 96-well plates (Becton-Dickinson) are coated with 2 μM of each experimental polypeptide in PBS. After overnight at incubation at 4°C, plates are washed twice in PBST (PBS containing 0.05% of Tween 20) and wells blocked with 1% BSA Fraction V in PBST. After 1 h incubation, plates are washed thrice in PBST and a range of dilutions of test and control sera in PBST added to the wells. After 2 h incubation, plates are washed six times in PBST, and primary anti-mouse-Ig2a sera are added to all wells. After 1 h incubation, plates are washed six times in PBST, and anti-primary anti-mouse-Ig2a sera added to all wells. After 1 h incubation, plates are washed seven times with in PBST and TMB substrate to all wells. After 20-30 minutes incubation, the reaction is stopped with HCl and the absorbance at 450 nm is read.

Statistical analysis

Statistically significant differences in the IFN-γ response to different antigens between the test and control groups are established through non-parametric Mann-Whitney analysis of the samples. Differences are considered statistically significant if the p value is below 0.05.

EXPERIMENT 3 - Assessment of the immune response to various polypeptides:

The following polypeptides were investigated: SEQ ID 20, SEQ ID 28, SEQ ID 30, SEQ ID 31, SEQ ID 32 and SEQ ID 35. These peptides were mixed together to form a candidate vaccine for testing (called the AGS peptide mix).

The type and level of the immune response induced by vaccination with these peptides was assessed according to the protocol shown below:

Day 1: Immunise 2 groups of 4 CD1 mice each with the following subcutaneous doses of candidate vaccine products:

- Non-relevant-peptide (NRP) mix (10 nmol each) + ISA-51
- AGS peptide mix (10 nmol each) + ISA-51

Day 15: Boost all animals with the same doses of candidate vaccine products.

Day 21: Terminally bleed all animals. Harvest spleens individually and test for IFN-gamma reactivity to:

- Individual AGS peptides (2 μM each)
- AGS mix (0.5 μM and 2 μM each)
- Con A (7.5 μg/ml)
- Blank

Post-Day 21: Test all sera for reactivity against AGS peptides.

Results

IFN-gamma production following 96 hour stimulation *in vitro* with the antigens is indicated in the graph in Figure 5.

IFN-gamma responses to SEQ ID 28, SEQ ID 30 and SEQ ID 35 as well as to the AGS-mix preparation are statistically significant ($p<0.05$).

SEQ ID 20, SEQ ID 31 and SEQ ID 32 induce a higher response in the AGS-mix immunised animals, but they also appear to be stimulatory, in a non specific fashion, to the splenocytes of NRP-mix immunised mice.

The total Ig response in sera to the antigens indicated in the graph in Figure 6. Total Ig responses to SEQ ID 20 and SEQ ID 30 are statistically significant ($p<0.05$).

EXPERIMENT 4 - Challenge study following AGS peptide mix immunisation:

To test the capacity of the AGS-mix preparation to confer protection against natural malaria infection, CD1 mice were immunised and challenged according to the protocol below:

Day 0: Test bleed all animals. Retain samples for further analysis.

Day 1: Immunise 2 (two) groups of 8 CD1 mice each with the following subcutaneous doses of candidate vaccine products:

- Non-relevant-peptide (NRP) mix (10 nmol each) + ISA-51
- AGS peptide mix (10 nmol each) + ISA-51

Day 7: Test bleed all animals. Retain samples for further analysis.

Day 14: Boost animals with the same doses of candidate vaccine products.

Day 21: Test bleed all animals. Retain samples for further analysis.

Day 28: All animals are challenged via bite of 8 infected mosquitoes (i.e. *Plasmodium yoelii nigeriensis* infected *Anopheles gambiae*) in the belly area. All animals are maintained until parasitemia is first established or for a maximum of 6 weeks after challenge with infected mosquitoes.

Day 70 (max): Terminally bleed all animals.

Results

Total Ig response in sera to the AGS-mix at day 21 is indicated in the graph in Figure 7.

One animal in the AGS-mix immunised group showed a significantly lower Total Ig response than the remaining animals in the group (<50% average total Ig response in the AGS-mix group).

On the day of challenge, one animal in the NRP-mix immunised group and two animals in the AGS-mix immunised group could not be challenged due to a shortage of infectious mosquitoes.

Of the animals that were challenged, those in the AGS-mix immunised group showed an increased survival rate than those in the control NRP-mix immunised group (see Figure 8). The one animal that died in the AGS-mix immunised group was the same one that had failed to develop a strong antibody response to the AGS preparation.

CLAIMS:

1. A polypeptide composition comprising one or more polypeptides, which polypeptides are immunogenic in a vertebrate such that they cause the vertebrate to produce immune system cells capable of recognising at least one epitope from an arthropod saliva protein fraction, wherein the arthropod saliva protein fraction has a mass of 40kDa or less, and wherein the polypeptides are selected independently from:

(a) the polypeptide sequences of SEQ ID 1-44 or sub-sequences from these sequences, the sub-sequences having 7 amino acids or more:

SEQ ID 1	HLTLFTVAVLLLAAAALLLLLPPAYSTTLTPP
SEQ ID 2	PLSYCHLFLTHTLARALSFSRSDCL
SEQ ID 3	KNVFFALLLVVLVCCLVSVQGNEI
SEQ ID 4	KLLVLLICLFFYHTHCTTAYLWLAMGV
SEQ ID 5	FLKGSFPRFQMCVMLIGFFSSAKCL
SEQ ID 6	NDYQALLGLCCPWIDLAAADLPMRRHAKA
SEQ ID 7	FYSVGKLVKVLLVMAVCCLLLCTAPTGADPL
SEQ ID 8	MKFAFAFVLIALFAVFAVSQALPQPEQAAA
SEQ ID 9	DGASAITKIVLELTPEQAAAV
SEQ ID 10	TLFIFLVCCQIPLFGIMSSDSADPFYWIRVILA
SEQ ID 11	GRVMCLLRLMSTLLVVLSIVGK
SEQ ID 12	LYSGYRLLVLLVMTVCCLLLFIAPTGADPLPGQTQRTL
SEQ ID 13	MYCVIKGKTGGYCNSEGLCTCRAEDLHFLLKPIINKD
SEQ ID 14	NAEDPRTELIGCGSVLFHLAANRLSLQLEEFAVCKR
SEQ ID 15	ALIGLLLCSVQSVTANDPVDALGACSGNLFGLLMTRL
SEQ ID 16	SKLFVLAFLCLALVVVVQSAPQYARGDVPT
SEQ ID 17	SMLVAFATLSVALVVVVAIPANFNYGGGGGYFINGTGQ
SEQ ID 18	IYEKLPAYLSEVSARVNVLQVSLQHDLPNLQ
SEQ ID 19	EMKLAKVALVTISLWFMAWTPYLVINFTGI
SEQ ID 20	LLPAKVIPDKTAAYVAYGGQETLVEHVEVLV
SEQ ID 21	FYTCFLGTSSLAGFKNAVDYDELLKAG

SEQ ID 22	VLEVLGFVEDNGELVFQELLGVLKMVDPDGD
SEQ ID 23	KLTPTVVVVLLCLTFVADALTIQELRAQIAQQRIQQRYGVTVATT
SEQ ID 24	SLSDYGLIELKEHCLECCQKDTEADSKLKVYPAAVLEV
SEQ ID 25	TYICFILHGVSEIIPQQQKKTMKFLLLVASVLCLVLI
SEQ ID 26	RYFVVIALICPLIIVETLAV
SEQ ID 27	LLLYLDAADLRRALHQYQLLAAQGDRHLPQQIVKFV
SEQ ID 28	VLLTPALQAYIMDEHNLNRSNIALGRIRPYPSAVKMP
SEQ ID 29	VLKGETHKALKLKDGGHYLVEFKSIYM
SEQ ID 30	VLHSMLVNASLAEMVKESYQTHGADGRMVVRMLKFVRLLP
SEQ ID 31	RVRALRALLETLLQHQGEQNNDVYLIRLAHET
SEQ ID 32	ELQQALSSLNAGSGSCAEVFNAYLPVHNKYIGVSRKI
SEQ ID 33	KFYRLISTLLVVVVIAPRHQCSPFFFQYNRPYL
SEQ ID 34	NYVPDVSALEQDIIEVDPETKEMLKHLDFNNIVVQL
SEQ ID 35	QYSMECLEAAEPKYLDGLKTLADETAQC
SEQ ID 36	EYAQVTKMLGNGRLEAMCFDGVKRLCHIRGKL
SEQ ID 37	KLFLTLLSTLSVAMVFALPAHHHSRG
SEQ ID 38	ELEEARLVAEELEERQQELDYLKRYLVGRLQAV
SEQ ID 39	SYFLTVCLLALVQSETVQD
SEQ ID 40	AMTNANLVGLTISLAYAIFFLLYTPPTGRSS
SEQ ID 41	SFAWLLYGIILRSNFLVVQNLMALALSAVQLSLFII
SEQ ID 42	AFPFISGFLSCFMWLKYGVLTEESTLILVNFIGSAL
SEQ ID 43	GLLCCCLAVLFFASPLTMLAHVIR
SEQ ID 44	LLLAMVLLPLLLLESVVPYAAAEKVW

(b) the sequences defined by the following amino acid residues of an arthropod saliva protein, or sub-sequences from these sequences, the sub-sequences having 7 amino acids or more:

residues 2-33 of >gi|18389913|gb|AAL68793.1|AF457563_1 hypothetical protein 16 [Anopheles gambiae]

residues 2-26 of >gi|18389909|gb|AAL68791.1|AF457561_1 hypothetical protein 14 [Anopheles gambiae]

residues 2-25 of >gi|18389907|gb|AAL68790.1|AF457560_1 hypothetical protein 13 [Anopheles gambiae]

residues 10-36 of >gi|18389903|gb|AAL68788.1|AF457558_1 hypothetical protein 11 [Anopheles gambiae]

residues 2-26 of >gi|62546227|gb|AAX86005.1| hyp3.5 precursor [Anopheles gambiae]

residues 14-42 of gi|18389899|gb|AAL68786.1|AF457556_1 salivary gland 7-like protein [Anopheles gambiae]

residues 3-33 of >gi|18389911|gb|AAL68792.1|AF457562_1 hypothetical protein 15 [Anopheles gambiae]

residues 1-30 of >gi|62546225|gb|AAX86004.1| hyp6.3 precursor [Anopheles gambiae]

residues 34-54 of >gi|62546225|gb|AAX86004.1| hyp6.3 precursor [Anopheles gambiae]

residues 38-70 of >gi|17026153|emb|CAD12038.1| Sec61 protein [Anopheles gambiae]

residues 2-23 of >gi|62546223|gb|AAX86003.1| hyp6.2 precursor [Anopheles gambiae]

residues 17-54 of >gi|18389915|gb|AAL68794.1|AF457564_1 hypothetical protein 17 [Anopheles gambiae]

residues 57-93 of >gi|87080391|gb|ABD18596.1| defensin [Anopheles gambiae]

residues 22-57 of >gi|18389901|gb|AAL68787.1|AF457557_1 hypothetical protein 10 [Anopheles gambiae]

residues 7-43 of >gi|18389905|gb|AAL68789.1|AF457559_1 hypothetical protein 12 [Anopheles gambiae]

residues 3-32 of >gi|4127344|emb|CAA76832.1| cE5 protein [Anopheles gambiae]

residues 3-40 of >gi|4210617|emb|CAA10259.1| SG2 protein [Anopheles gambiae]

residues 91-121 of >gi|4127309|emb|CAA76820.1| hypothetical protein [Anopheles gambiae]

residues 65-94 of >gi|4375824|emb|CAA76825.1| opsin [Anopheles gambiae]

residues 41-71 of >gi|62546233|gb|AAX86008.1| unknown [Anopheles gambiae]

residues 117-143 of >gi|3378531|emb|CAA03872.1| D7r2 protein [Anopheles gambiae]

residues 63-93 of >gi|3378529|emb|CAA03871.1| D7r3 protein [Anopheles gambiae]

residues 23-67 of >gi|18389893|gb|AAL68783.1|AF457553_1 mucin-like protein [Anopheles gambiae]

residues 43-80 of >gi|18389881|gb|AAL68777.1|AF457547_1 selenoprotein [Anopheles gambiae]

residues 6-42 of >gi|18389879|gb|AAL68776.1|AF457546_1 30 kDa protein [Anopheles gambiae]

residues 4-23 of >gi|18378603|gb|AAL68639.1|AF458073_1 D7-related 5 protein [Anopheles gambiae]

residues 20-55 of >gi|18389897|gb|AAL68785.1|AF457555_1 salivary gland 1-like 4 protein [Anopheles gambiae]

residues 59-95 of >gi|18389883|gb|AAL68778.1|AF457548_1 antigen 5-related 1 protein [Anopheles gambiae]

residues 158-184 of >gi|83016748|dbj|BAE53441.1| DsRed [synthetic construct]

residues 37-76 of >gi|18389895|gb|AAL68784.1|AF457554_1 salivary gland 1-like 3 protein [Anopheles gambiae]

residues 191-222 of >gi|18389895|gb|AAL68784.1|AF457554_1 salivary gland 1-like 3 protein [Anopheles gambiae]

residues 113-149 of >gi|18389891|gb|AAL68782.1|AF457552_1 D7 protein long form [Anopheles gambiae]

residues 1-37 of >emb|CAC35527.1| gSG9 protein [Anopheles gambiae]

residues 81-120 of >sp|Q9U9L1|RS17_ANOGA 40S ribosomal protein S17

residues 111 to 142 of >emb|CAC35523.1| gSG7 protein [Anopheles gambiae]

residues 32-67 of >gb|AAD47075.1|AF164151_1 translation initiation factor 4C (1A) [Anopheles gambiae]

residues 1-29 of >emb|CAC35519.1| gSG2-like protein [Anopheles gambiae]

residues 106-142 of >emb|CAC35451.1| hypothetical protein [Anopheles gambiae]

residues 6-28 of > emb|CAC35524.1| D7r4 protein [Anopheles gambiae]

residues 70-104 of >ref|XP_001230998.1| ENSANGP00000014906 [Anopheles gambiae str. PEST]

residues 174-213 of >ref|XP_316361.2| ENSANGP00000012984 [Anopheles gambiae str. PEST]

residues 41-80 of >ref|XP_314140.3| ENSANGP00000015780 [Anopheles gambiae str. PEST]

residues 126-153 of >ref|XP_314140.3| ENSANGP00000015780 [Anopheles gambiae str. PEST]

residues 5-34 of >emb|CAC35522.1| gSG6 protein [Anopheles gambiae]

(c) or from polypeptide sequences having 85% homology or more with one or more of the sequences in (a) or (b) and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

2. A polypeptide composition according to claim 1, wherein the polypeptides are selected independently from SEQ ID 1-6, 20, 28, 30-32 and 35, or sub-sequences from these sequences, the sub-sequences having 7 amino acids or more, or from polypeptide sequences having 85% homology or more with one of these sequences and contained in one or more of the following databases: GenBank, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), or CDS translations of these.

3. A polypeptide composition according to claim 1 or claim 2, which composition comprises 2 or more polypeptides, preferably from 2 to 12 polypeptides, or more preferably from 2-6 polypeptides.

4. A polypeptide composition according to any preceding claim, wherein the arthropod saliva protein fraction has a mass of from 20kDa to 40kDa, or has a mass of 20kDa or less.

5. A polypeptide composition according to any preceding claim, wherein the one or more polypeptides are cytotoxic T lymphocyte (CTL) epitopes, and/or the one or more polypeptides are T helper lymphocyte (Th) epitopes, and/or the one or more polypeptides are B lymphocyte epitopes.

6. A polypeptide composition according to any preceding claim, comprising 2, 3, 4, 5, 6 or more epitopes.

7. A polypeptide according to any preceding claim, which is immunogenic against a plurality of arthropod salivary proteins.

8. A polypeptide composition according to any preceding claim, wherein the arthropod saliva protein fraction is a mosquito saliva protein fraction from any mosquito species, preferably from anopheles gambiae.

9. A polypeptide composition according to any preceding claim, further comprising one or more further sequences from an arthropod salivary protein.

10. A polypeptide composition according to any preceding claim, wherein the sub-sequence comprises 8, 9, 10, or 11 amino acids or more.

11. A polypeptide composition according to any preceding claim, wherein each polypeptide comprises no more than 100 amino acids, preferably from 15-75 amino acid residues, and more preferably from 17-55 amino acid residues.

12. A polypeptide composition according to any preceding claim, consisting substantially of one ore more polypeptides having a sequence from the sequences SEQ ID 1-44.

13. A polypeptide composition according to any preceding claim, in which polypeptides having sequences of SEQ ID 1-6 are all present, and/or in which polypeptides having sequences of SEQ ID 20, 28, 30-32 and 35 are all present.

14. A polypeptide composition, as defined in any preceding claim for use in medicine.

15. A polypeptide composition, as defined in any preceding claim for veterinary use.

16. A method of producing a polypeptide composition as defined in any preceding claim, which method comprises mixing or combining one or more polypeptides as defined in any preceding claim with one or more further polypeptides as defined in any preceding claim, and/or with one or more further components, such as a carrier, an excipient, an adjuvant, a buffer, or a stabiliser.

17. A polypeptide construct, which construct comprises a polypeptide composition as defined in any of claims 1-15 and a carrier.

18. A method of producing a polypeptide construct as defined in claim 17, which method comprises attaching, combining or mixing a polypeptide composition as defined in any of claims 1-15 with the carrier.

19. A polypeptide construct or a method according to claim 17 or claim 18, wherein the carrier comprises an adjuvant and/or an excipient.

20. A nucleic acid construct comprising one or more nucleic acids comprising the coding sequence of the one or more polypeptides in the polypeptide composition as defined in any of claims 1-15.

21. A nucleic acid construct according to claim 20, which is a DNA or RNA construct.

22. A nucleic acid construct according to claim 20 or claim 21, which comprises a plasmid or a recombinant virus.

23. A medicament or vaccine composition against an arthropod borne disease, comprising a polypeptide composition, polypeptide construct, or nucleic acid construct as defined in any of claims 1-15, 17, and 19-22 and optionally an appropriate excipient and/or adjuvant.

24. A method of producing a medicament or vaccine composition as defined in claim 23, which method comprises attaching, combining or mixing a polypeptide composition, a polypeptide construct and/or a nucleic acid construct, as defined in any of claims 1-15, 17 and 19-22 with an appropriate excipient and/or adjuvant.

25. Use of a polypeptide composition, a polypeptide construct, a nucleic acid construct, or a medicament or vaccine composition as defined in any of claims 1-15, 17 and 19-23 in the manufacture of a medicament or vaccine, effective in the treatment or prevention of an arthropod borne disease.

26. A method of treating or preventing an arthropod borne disease, which method comprises administering a polypeptide composition, a polypeptide construct, a nucleic acid construct, or a medicament or vaccine composition, as defined in any of claims -15, 17 and 19-23 to a vertebrate.

27. A method according to claim 26, wherein the vertebrate is selected from a mammal, a bird, a reptile and a fish, and preferably from a human, a domestic animal, or a farm animal, a bovine animal, or a fowl.

28. A polypeptide composition, polypeptide construct, nucleic acid construct, method, medicament, vaccine or use according to any preceding claim, wherein the arthropod is a hematophagous arthropod.

29. A polypeptide composition, polypeptide construct, nucleic acid construct, method, medicament, vaccine or use according to claim 28, wherein the hematophagous arthropod is selected from mosquitoes, ticks, sandflies, mites, and fleas.

30. A medicament, vaccine composition, method, or use according to any of claims 23-29, wherein the arthropod borne disease is a disease selected from malaria, filariasis, trypanosomiasis, leishmaniasis, onchocerciasis, dengue, lyme disease, tularaemia, relapsing fever, typhus, plague, yellow fever, rift valley fever, and encephalitis, such as St. Louis encephalitis, west Nile fever, eastern equine encephalitis, La Crosse encephalitis and eastern equine encephalitis.

31. A polypeptide composition, polypeptide construct, nucleic acid construct, method, medicament, vaccine or use according to any preceding claim, wherein the arthropod is selected from *Anopheles sp., Aedes sp., Culex sp., Mansonia sp., Chrysops sp., Simulium sp., Triatoma sp., Rhodnius sp., Panstrongylus sp., Glossina sp., Tabanus sp., Phlebotomus sp., Lutzomyia sp., Pediculus sp, Ornithodoros sp., Ixodes sp., Pediculus sp., Pulex sp., Hemagogous spegazzani., Ochlerotatus sp.,* and *Coquillettidia sp.*

FIG. 1

IEF Gel of Anopheles gambiae
Salivary Glands
(Comassie Blue Stain)

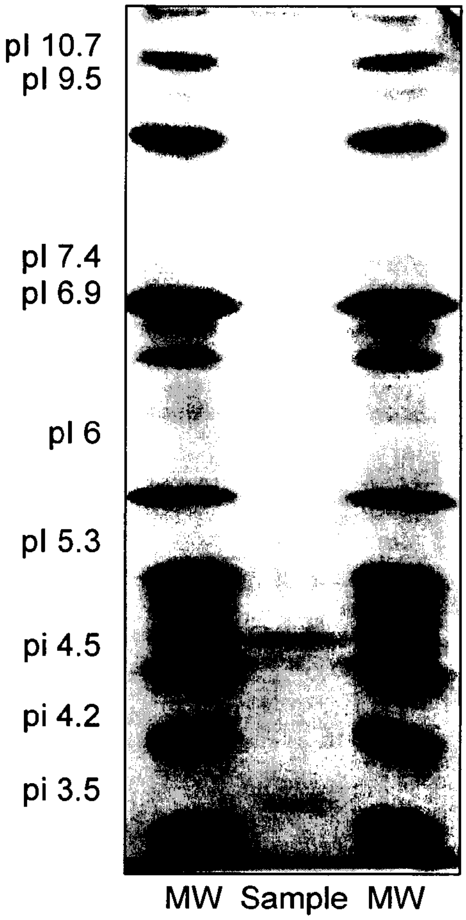


FIG. 2

SDS-PAGE of IEF gel of Anopheles gambiae Salivary Glands (silver stain)
Red rectangle indicates location of targeted proteins (<30 kDa)

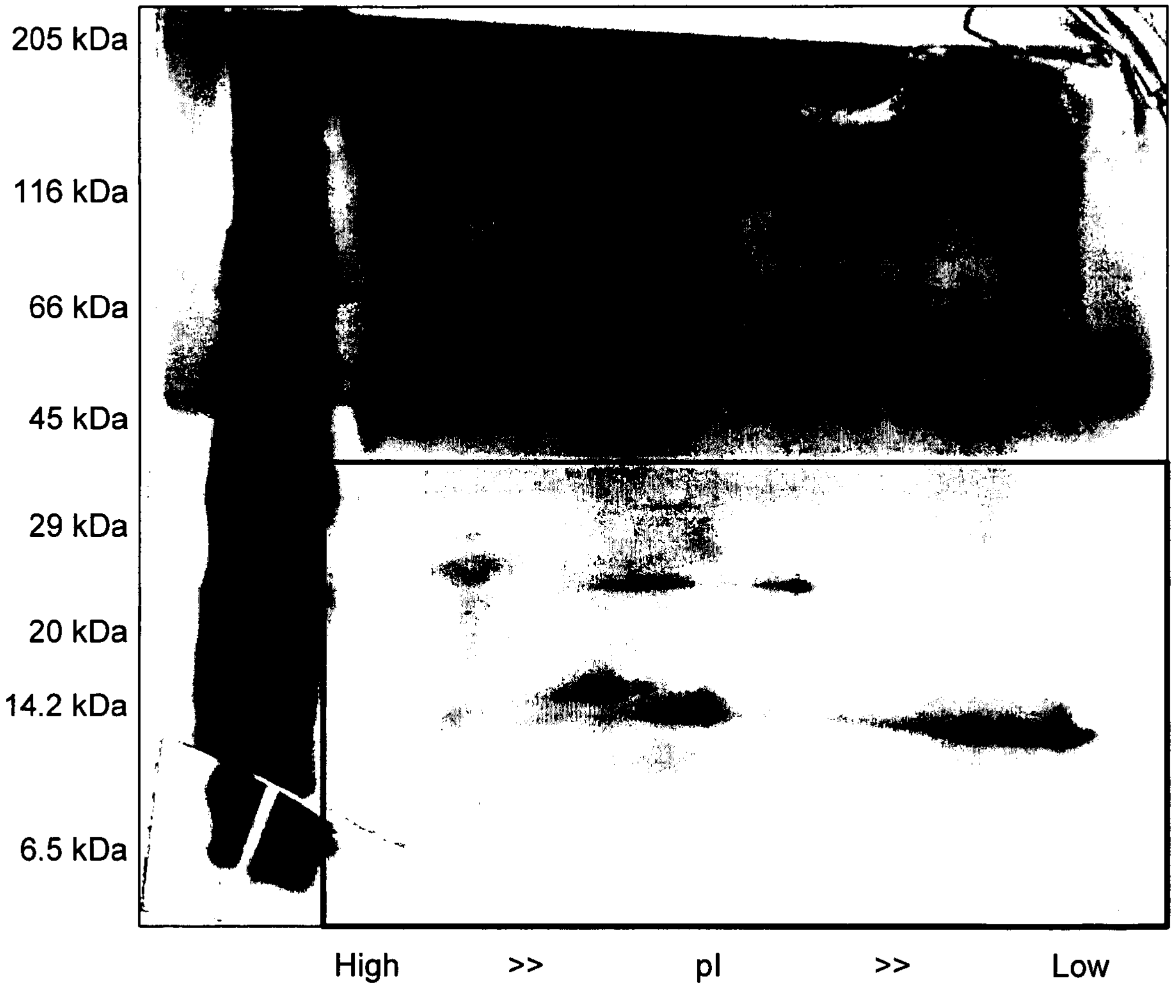


FIG. 3

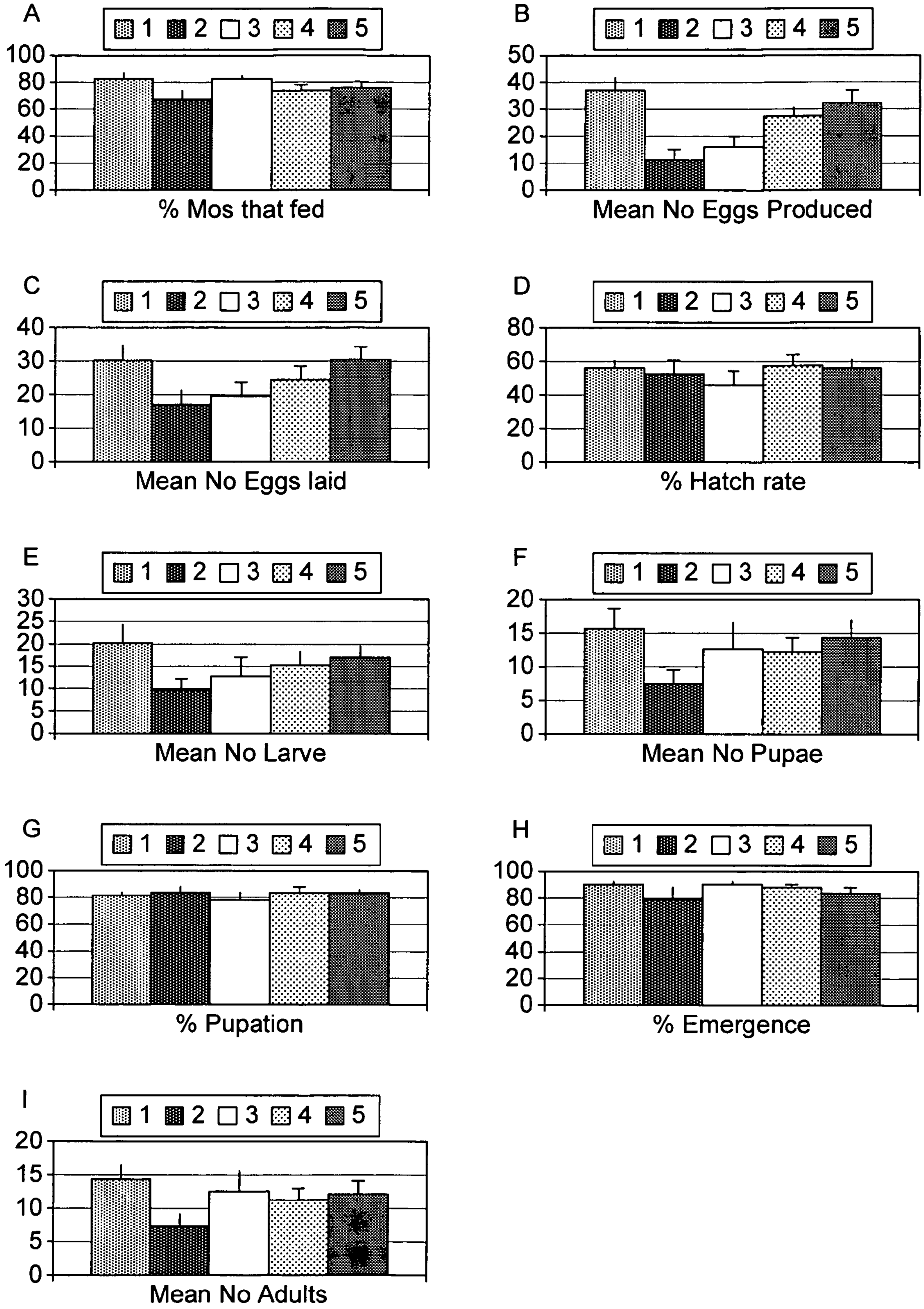

A
▨ 1 ▨ 2 ☐ 3 ▨ 4 ▨ 5
100
80
60
40
20
0
% Mos that fed
B
▨ 1 ▨ 2 ☐ 3 ▨ 4 ▨ 5
50
40
30
20
10
0
Mean No Eggs Produced
C
▨ 1 ▨ 2 ☐ 3 ▨ 4 ▨ 5
40
30
20
10
0
Mean No Eggs laid
D
▨ 1 ▨ 2 ☐ 3 ▨ 4 ▨ 5
80
60
40
20
0
% Hatch rate
E
▨ 1 ▨ 2 ☐ 3 ▨ 4 ▨ 5
30
25
20
15
10
5
0
Mean No Larve
F
▨ 1 ▨ 2 ☐ 3 ▨ 4 ▨ 5
20
15
10
5
0
Mean No Pupae
G
▨ 1 ▨ 2 ☐ 3 ▨ 4 ▨ 5
100
80
60
40
20
0
% Pupation
H
▨ 1 ▨ 2 ☐ 3 ▨ 4 ▨ 5
100
80
60
40
20
0
% Emergence
I
▨ 1 ▨ 2 ☐ 3 ▨ 4 ▨ 5
20
15
10
5
0
Mean No Adults

FIG. 4

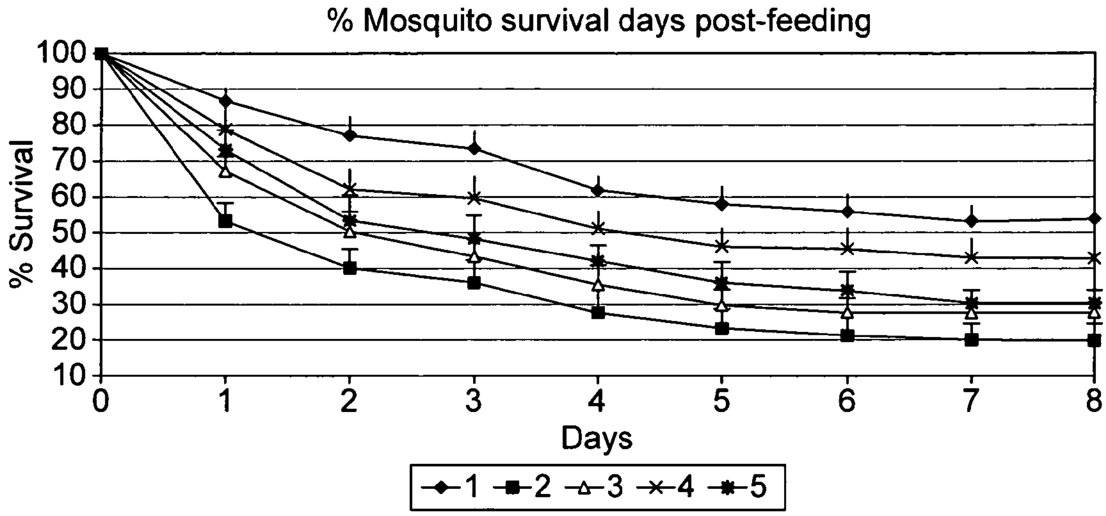

% Mosquito survival days post-feeding
% Survival
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
Days
1
2
3
4
5

FIG. 6

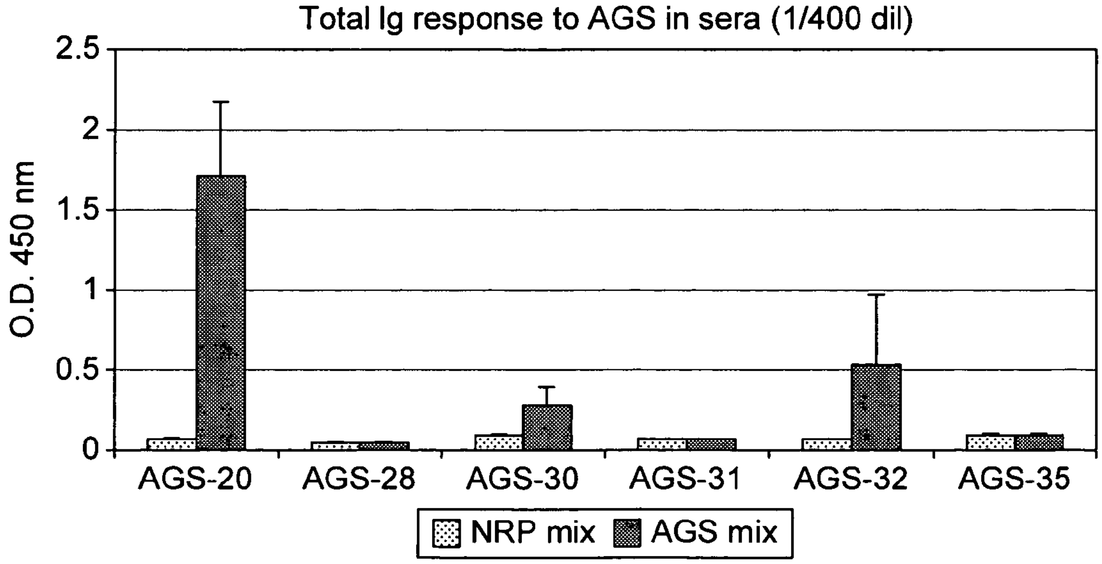

Total Ig response to AGS in sera (1/400 dil)
O.D. 450 nm
2.5
2
1.5
1
0.5
0
AGS-20
AGS-28
AGS-30
AGS-31
AGS-32
AGS-35
NRP mix
AGS mix

FIG. 5

To Blank

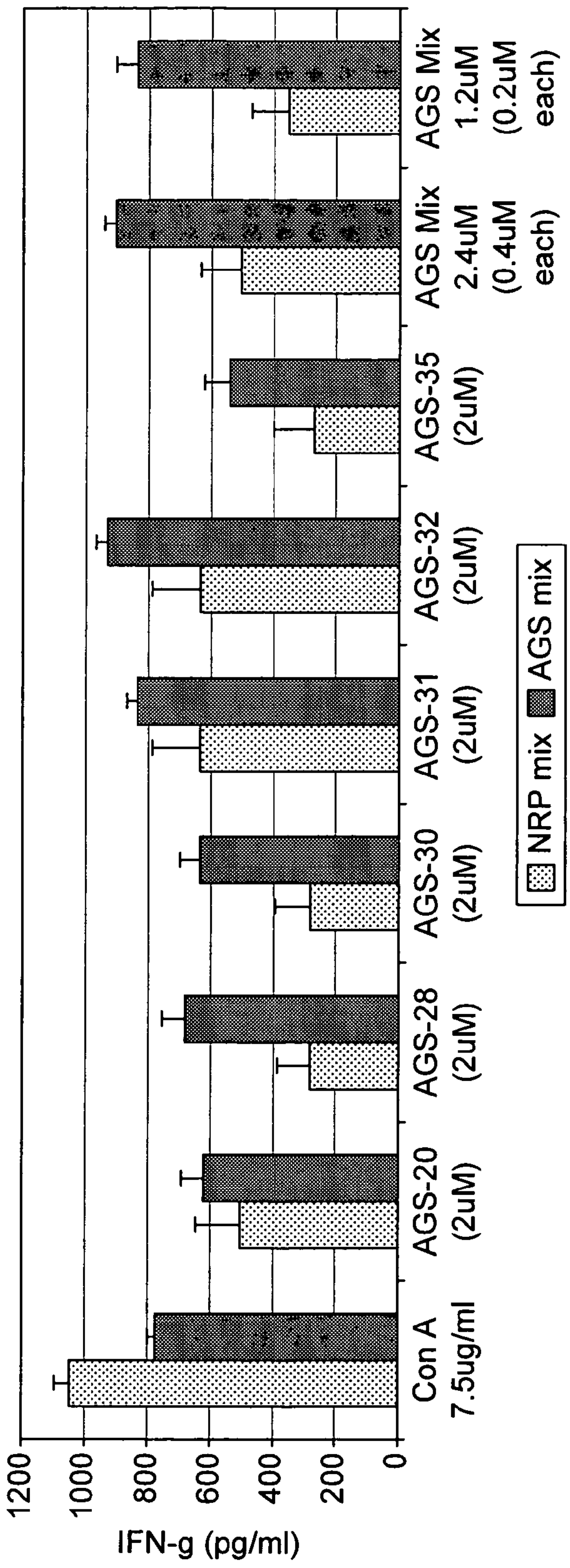


FIG. 7

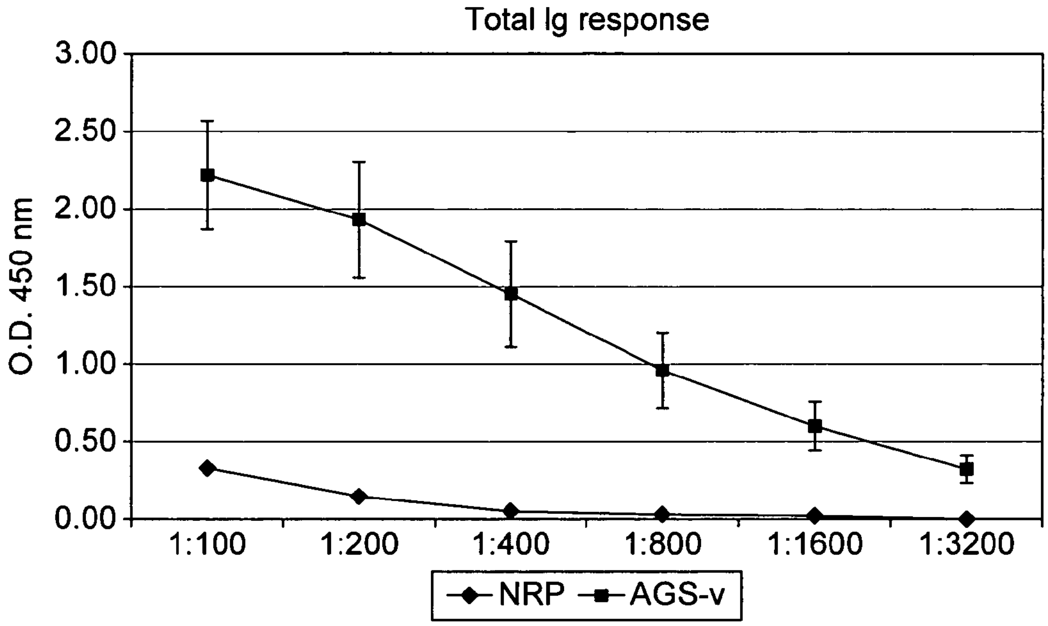

Total Ig response
O.D. 450 nm
3.00
2.50
2.00
1.50
1.00
0.50
0.00
1:100
1:200
1:400
1:800
1:1600
1:3200
NRP
AGS-v

FIG. 8

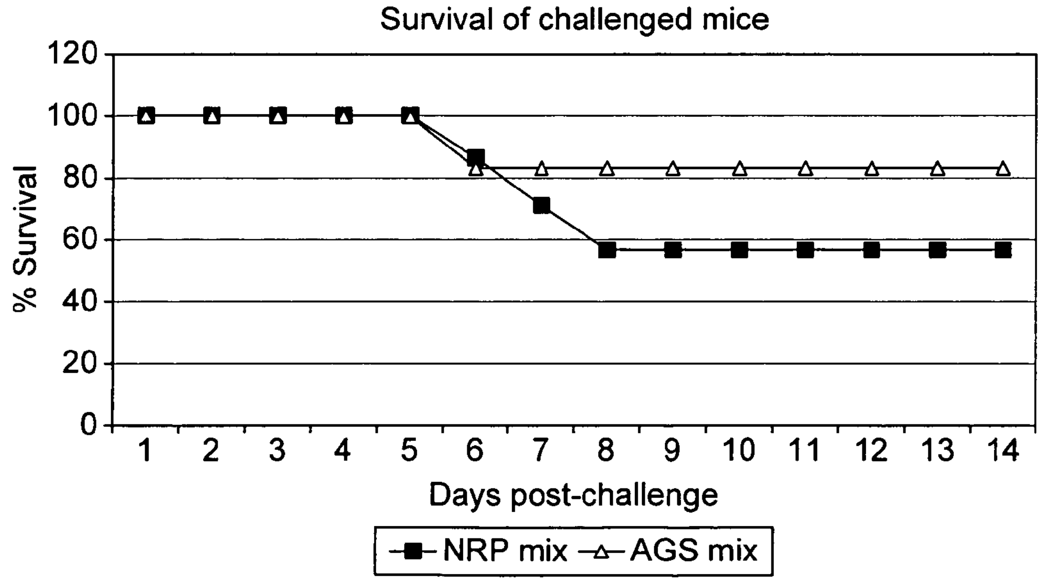

Survival of challenged mice
% Survival
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Days post-challenge
NRP mix
AGS mix

FIG. 9


97 kDa
68 kDa
43 kDa
29 kDa
18 kDa
14 kDa
MW Standards
An. Gambiae SGP