# PEPTIDES FOR VACCINE

## Field of the Invention

The present invention relates to compositions comprising peptides for preventing or treating allergy to house dust mites, and in particular to optimal combinations of peptides for preventing or treating said allergy.

## Background of the Invention

T-cell antigen recognition requires antigen presenting cells (APCs) to present antigen fragments (peptides) on their cell surface in association with molecules of the major histocompatibility complex (MHC). T cells use their antigen specific T-cell receptors (TCRs) to recognise the antigen fragments presented by the APC. Such recognition acts as a trigger to the immune system to generate a range of responses to eradicate the antigen which has been recognised.

Recognition of external antigens by the immune system of an organism, such as man, can in some cases result in diseases, known as atopic conditions. Examples of the latter are the allergic diseases including asthma, atopic dermatitis and allergic rhinitis. In this group of diseases, B lymphocytes generate antibodies of the IgE class (in humans) which bind externally derived antigens, which are referred to in this context as allergens since these molecules elicit an allergic response. Production of allergen-specific IgE is dependent upon T lymphocytes which are also activated by (are specific for) the allergen. Allergen-specific IgE antibodies bind to the surface of cells such as basophils and mast cells by virtue of the expression by these cells of surface receptors for IgE.

Crosslinking of surface bound IgE molecules by allergen results in degranulation of these effector cells causing release of inflammatory mediators such as histamine, 5-hydroxtryptamine and lipid mediators such as the sulphidoleukotrienes. In addition to IgE-dependent events, certain allergic diseases such as asthma are characterised by IgE-independent events.

Allergic IgE-mediated diseases are currently treated with agents which

provide symptomatic relief or prevention. Examples of such agents are anti-histamines, β2 agonists, and glucocorticosteroids. In addition, some IgE-mediated diseases are treated by desensitisation procedures that involve the periodic injection of allergen components or extracts. Desensitisation treatments may induce an IgG response that competes with IgE for allergen, or they may induce specific suppressor T cells that block the synthesis of IgE directed against allergen. This form of treatment is not always effective and poses the risk of provoking serious side effects, particularly general anaphylactic shock. This can be fatal unless recognised immediately and treated with adrenaline. A therapeutic treatment that would decrease or eliminate the unwanted allergic-immune response to a particular allergen, without altering the immune reactivity to other foreign antigens or triggering an allergic response itself would be of great benefit to allergic individuals.

House dust mites are universally recognised as a major cause of allergic diseases in humans and animals, including asthma, allergic rhinitis and allergic dermatitis. Two closely related species of mite are responsible for the majority of house dust mite allergy worldwide. These are *Dermatophagoides pteronyssinus* (predominantly in Europe) and *Dermatophagoides farinae* (predominantly in America). House dust mite allergens are mainly derived from proteins from the lining of the mite gut, which are present in the faeces, and are typically referred to as Der p (for *D. pteronyssinus*) or Der f (for *D. farinae*) proteins. An average mite will produce approximately 20 faecal pellets each day of its life: twice its own body weight. One gram of dust can typically contain up to 500 mites, while a mattress can hold more than two million. The amount of mite material present increases with age. One tenth of the weight of a six-year old pillow can consist of mites and mite debris. In a carpet, there will typically be between 1,000 and 10,000 mites per square metre.

Allergic diseases, particularly asthma, are a huge and expanding problem in the industrialised nations of the world. It has been calculated that 5-10% of the population of the major industrialised nations suffers from asthma. Of those, approximately one fifth will have severe asthma requiring frequent hospitalisation. The cost of asthma within the United States has been calculated as $12.6 billion (£7.9 billion) per year. Figures for Europe are even higher. A Canadian study estimated the

costs of asthma as averaging £21 per year for every member of the population of the major industrialised nations. 2,000 people every year will die as a result of asthma in the United Kingdom alone.

Asthma is a chronic disease caused by allergic reactions and irritation within the respiratory system. Between 50% and 90% of asthmatics who react to airborne material are sensitive to dust mite allergens, and in one British study 10% of the general population reacted to dust mite allergens. Almost two hundred million Americans live in areas severely affected by house dust mite infestation. Sensitisation to this material occurs in childhood, mainly between three and six months of age but asthma is lifelong.

A therapeutic or preventative treatment would therefore be of great benefit to humans that suffer or are at risk of suffering from house dust mite allergy.

## Summary of the Invention

The present inventors have discovered that certain combinations of peptide fragments derived from the Group 1 dust mite allergen (Der p 1, Der f 1), Group 2 dust mite allergen (Der p 2, Der f 2) and Group 3 dust mite allergen (Der p 7, Der f 7) are particularly useful in desensitising individuals to these allergens. The polypeptide combinations of the invention have been selected for their ability to induce a cytokine response in a high proportion of subjects from a panel of house dust mite allergic individuals.

The polypeptides of the invention were initially selected as T cell epitopes through use of both *in silico* and *in vitro* assessments of peptide – MHC binding characteristics. See for example Table 3 which demonstrates the ability of a range of peptides derived from the above allergens to bind to multiple DR types in MHC class II binding assays. Additional epitopes were identified by homology. These candidate polypeptides were then further screened for potential use in tolerisation.

A difficulty associated with approaches to desensitisation based on peptide immunisation lies in how to select an appropriate size and region of the allergen as the basis for the peptide to be used for immunisation. The size of the peptide of

choice is crucial. If the peptide is too small, the vaccine would not be effective in inducing an immunological response. If the peptides are too large, or if the whole antigen is introduced into an individual, there is the risk of inducing adverse reactions, such as anaphylaxis, which may be fatal.

The polypeptides of the invention have been selected to retain T cell specificity whilst being small enough in size to not possess significant tertiary structure that would enable them to retain the conformation of an IgE-binding epitope of the whole molecule. The polypeptides of the invention therefore do not induce significant crosslinking of adjacent specific IgE molecules on cells such as mast cells and basophils and consequently do not cause significant histamine release.

An advantage of the invention is the ability of the peptides to broadly target Major Histocompatibility Complex (MHC) molecules. T cell receptors (TCRs) are highly variable in their specificity. Variability is generated, as with antibody molecules, through gene recombination events within the cell. TCRs recognise antigen in the form of short peptides bound to molecules encoded by the genes of the Major Histocompatibility Complex (MHC). These gene products are the same molecules that give rise to "tissue types" used in transplantation and are also referred to as Human Leukocyte Antigen molecules (HLAs) which terms may be used interchangeably. Individual MHC molecules possess peptide binding grooves which, due to their shape and charge are only capable of binding a limited group of peptides. The peptides bound by one MHC molecule may not necessarily be bound by other MHC molecules.

When a protein molecule such as an antigen or allergen is taken up by antigen presenting cells such as B lymphocytes, dendritic cells, monocytes and macrophages, the molecule is enzymatically degraded within the cell. The process of degradation gives rise to peptide fragments of the molecule which, if they are of the appropriate size, charge and shape, may then bind within the peptide binding groove of certain MHC molecules and be subsequently displayed upon the surface of antigen presenting cells. If the peptide/MHC complexes are present upon the antigen presenting cell surface in sufficient numbers they may then activate T cells which bear the appropriate peptide/MHC-specific T cell receptors.

Due to the polymorphic nature of the MHC, individuals in an outbred population such as man will express different combinations of MHC molecules on their cell surfaces. Since different MHC molecules can bind different peptides from the same molecule based on the size, charge and shape of the peptide, different individuals will display a different repertoire of peptides bound to their MHC molecules. Identification of universal MHC-binding peptide epitopes in an outbred population such as man is more difficult than in inbred animals (such as certain strains of laboratory mice). On the basis of differential MHC expression between individuals and the inherent differences in peptide binding and presentation which this brings, it is unlikely that a single peptide can be identified which will be of use for desensitisation therapy in man.

The peptide combinations of the invention, however, provide a broad coverage of efficacy over the human population by targeting multiple different MHC molecules. A vaccine formulated with the peptides of the invention would therefore have broad utility.

The inventors' work has produced peptide combinations with the following characteristics:

- the combination induces a cytokine response in a high proportion of subjects from a panel of house dust mite allergic individuals
- the peptides of the combinations are soluble.

Accordingly, the present invention provides a composition for use in preventing or treating allergy to house dust mites by tolerisation comprising at least one polypeptide selected from HDM203B (SEQ ID 83), HDM201 (SEQ ID 80), HDM205 (SEQ ID 85), HDM203A (SEQ ID 82), HDM202 (SEQ ID 81), SEQ ID NO's 1 to 79, 84, or 86 to 104 (that is any one of SEQ ID NO's. 1 to 104) or a variant thereof. Typically, the composition comprises at least four polypeptides, wherein the polypeptides are independently selected from any of the following:

(i) a polypeptide of SEQ ID NO's 1 to 104; or

(ii) a variant of a polypeptide according to (i), wherein said variant is a polypeptide of length 9 to 30 amino acids that comprises a region consisting of:

- any of the sequences of (i); or

- a sequence which has at least 65% homology to any of the sequences of (i) which sequence is capable of tolerising an individual to any of the sequences of (i); or

(iii) a variant of a polypeptide according to (i), wherein said variant is a polypeptide of length 9 to 30 amino acids that comprises a region consisting of a sequence that represents either:

- a fragment of any of the sequences of (i); or
- a homologue of a fragment of any of the sequences of (i),

which sequence is capable of tolerising an individual to any of the sequences of (i) and has a length of at least 9 amino acids, and wherein said homologue has at least 65% homology to any 9 contiguous amino acids in any of the sequences of (i).

Description of the drawings

Figure 1 – Sequence comparison of Der p 1 versus Der f 1 (Fig 1A), Der p 2 versus Der f 2 (Fig 1B) and Der p 7 versus Der f 7 (Fig 1C). Regions containing epitopes are highlighted in grey. Locations of specific peptides of the invention are indicated by lines above or below the sequence. The sequence of Der p 1 is the publically available sequence with NCBI Accession No. P08176. The corresponding sequences for Der p 2 and Der p 7 (Table 6) are NCBI Accession Nos. P49278 and P49273, respectively. The sequence for Der f 1 is taken from NCBI Accession No. P16311, Der f 2 is from NCBI Accession No. Q00855 and Der f 7 is from NCBI Accession No. Q26456.

Figure 2 shows the percentage of individuals responsive to different peptides of the invention measured by production of IL13 or IFN-gamma.

Figures 3 and 4 show the percentage of individuals responsive to different peptide combinations of the invention measured by production of IL13 or IFN-gamma.

Description of the sequences mentioned herein

SEQ ID NOS: 1 to 104 provide the polypeptide sequences of the invention as set out in Tables 3 to 8. SEQ ID NOS. 105 onwards provide additional sequences.

Detailed description of the invention

The invention concerns peptides and combinations of peptides which can be used in tolerisation. Such peptides may comprise, consist of, or consist essentially of the sequences shown in any of HDM203B (SEQ ID 83), HDM201 (SEQ ID 80), HDM205 (SEQ ID 85), HDM203A (SEQ ID 82), HDM202 (SEQ ID 81), SEQ ID NO's 1 to 79, 84, or 86 to 104 (that is any one of SEQ ID NO's. 1 to 104). Variants of these specific peptides may also be used. The variants may comprise, consist of, or consist essentially of sequences which are fragments of either any of SEQ ID NO's 1 to 104 or homologues of any of SEQ ID NO's 1 to 104.

In one embodiment the invention relates to a composition for use in preventing or treating allergy to house dust mites. The composition typically comprises or consists at least four, five, six, seven, eight, nine, ten, eleven, or twelve polypeptides, up to a maximum of thirteen. In other words, the composition comprises between four and thirteen polypeptides. The polypeptides are independently selected from any of the following:

(i) a polypeptide of SEQ ID NO's 1 to 104; or

(ii) a variant of a polypeptide according to (i), wherein said variant is a polypeptide of length 9 to 30 amino acids that comprises a region consisting of:

- any of the sequences of (i); or
- a sequence which has at least 65% homology to any of the sequences of (i) which sequence is capable of tolerising an individual to any of the sequences of (i), or

(iii) a variant of a polypeptide according to (i), wherein said variant is a polypeptide of length 9 to 30 amino acids that comprises a region consisting of a sequence that represents either:

- a fragment of any of the sequences of (i); or
- a homologue of a fragment of any of the sequences of (i),

which sequence is capable of tolerising an individual to any of the sequences of (i) and has a length of at least 9 amino acids, and wherein said homologue has at least 65% homology to any 9 contiguous amino acids in any of the sequences of (i).

The invention also provides products and formulations comprising the polypeptides of the invention and compositions, products and vectors comprising polynucleotides capable of expressing the polypeptides of the invention for use in preventing or treating house dust mite allergy by tolerisation. Such tolerisation will typically be to an epitope (for example a MHC class II epitope) present in any of SEQ ID NO's 1 to 104.

*Peptide fragments of Group 1, Group 2 and Group 7 dust mite allergens*

The major allergens of the House dust mite include the Group 1 dust mite allergen (Der p 1, Der f 1), Group 2 dust mite allergen (Der p 2, Der f 2) and Group 3 dust mite allergen (Der p 7, Der f 7), wherein Der p "X" and Der f "X" indicate that the protein "X" is a homologue deriving from *D. pteronyssinus* and *D. farinae* respectively. As shown in Figure 1, each of the Der p proteins is highly homologous to its corresponding Der f protein.

The regions comprising MHC Class II-binding T cell epitopes are particularly highly conserved between the Der p and Der f homologues of a given protein. Peptides derived from the relevant regions of for example, protein 1 of either *D. pteronyssinus* or *D. farinae* are therefore suitable for use in preventing or treating house dust mite allergy by tolerisation to the Group 1 dust mite allergen. Similarly peptides derived from the relevant regions of protein 2 from either species are suitable for use in preventing or treating house dust mite allergy by tolerisation to the Group 2 dust mite allergen, and peptides derived from the relevant regions of protein 7 from either species are suitable for use in preventing or treating house dust mite allergy by tolerisation to the Group 7 dust mite allergen.

The Group 1 allergen is a cysteine protease homologous to papain. This enzyme has been found to cleave occludin, a protein component of intercellular tight junctions. This reveals one possible reason for the allergenicity of certain enzymes.

By destroying the integrity of the tight junctions between epithelial cells, Der p 1 and Der f 1 may gain abnormal access to subepithelial antigen-presenting cells, resident mast cells, and eosinophils.

The function of the Group 2 allergen is not known, although Der p 2 and Der f 2 show distant homology to a family of lipid-binding proteins. Serum IgE levels in response to stimulation with Der p 2 *in vivo* have been shown to represent approximately one third of the total serum IgE response to stimulation with whole mite extracts.

The function of the Group 7 allergen is also not known. Serum IgE levels in response to stimulation with Der p 7 *in vivo* have been shown to represent approximately one fifth of the total serum IgE response to stimulation with whole mite extracts.

The peptides of the invention are derived from the Group 1, Group 2 and Group 3 dust mite allergens as shown in Tables 3 to 8. The terms "peptide" and "polypeptide" are used interchangeably herein. The above proteins are also referred to herein as "the allergens".

Tables 3 to 8 set out the sequences of the peptides of the invention, indicating the parent protein from which each peptide derives. The sequences in Tables 4 to 6 are arranged in pairs. In each pair the upper sequence has been selected as a T cell epitope through use of peptide–MHC binding assays. The lower sequence has been selected by a homology search within the sequence of the alternative protein in the given dust mite allergen Group. For example, peptide HDM01 in Table 4 derives from Der p 1, the homologous sequence below it derives from Der f 1.

*Peptide combinations*

The composition typically comprises a combination of at least four different polypeptides of the invention, up to a maximum of thirteen different polypeptides. Accordingly, the composition of the invention may consist of four, five, six, seven, eight, nine, ten, eleven, twelve or thirteen peptides.

The composition of the invention may typically comprises at least one polypeptide or variant thereof (for example a functional variant) selected from a peptide which

derives from each of Der p 1, Der p 2 and Der p 7 (or the Der f equivalents). The polypeptide combinations in the composition of the invention are selected to provide as broad a coverage of the human population as possible, i.e. the composition of the invention will produce an immune response in a high proportion of dust mite allergic individuals, preferably more than 30%, 40%, 45%, 50%, 60% or 70% of dust mite allergic individuals in a panel or population of such individuals. The number of individuals in a population of dust mite allergic individuals may be any suitable number, typically at least 20, 30, 40, 50, 60, 70, 80, or at least 100 individuals. Preferably the population has MHC allele frequencies within the range of frequencies that are representative of the Caucasian population. Reference population allele frequencies for 11 common DRB1 allele families are shown in Table 1 (Data from HLA Facts Book, Parham and Barber).

The composition of the invention typically comprises at least one polypeptide selected from a polypeptide of HDM203B (SEQ ID 83), HDM202 (SEQ ID 81), HDM201 (SEQ ID 80), HDM205 (SEQ ID 85), HDM203A (SEQ ID 82), or a variant thereof. The composition preferably comprises at least two, three or four polypeptides independently selected from a polypeptide of HDM203B (SEQ ID 83), HDM202 (SEQ ID 81), HDM201 (SEQ ID 80), HDM205 (SEQ ID 85), HDM203A (SEQ ID 82), or a variant thereof, with the proviso that no more than one polypeptide or variant of SEQ ID NOS: 82 and 83 is selected.

Particular variants of HDM202 (SEQ ID 81) are HDM202D (SEQ ID 102; FKNRFLMSAEA), HDM202E (SEQ ID 103; FKNRFLMSAE) and HDM202H (SEQ ID 104; EFKNRFLMSAE), which are truncations of the HDM202 sequence. It is envisaged that each of these sequences can be modified to add at least one (and upto 6) residues at the N and/or C terminus selected from R, K, H, E and D.

Optionally, the composition may additionally comprise at least one additional polypeptide selected from a polypeptide of any of SEQ ID NOS: 5, 51, 52, 100, 101, 72, 73, 74, or a variant thereof. The at least one additional polypeptide is preferably a polypeptide of any of SEQ ID NOS: 51, 73, 100 and 101.

Optionally, the composition may additionally comprise at least one additional polypeptide selected from a polypeptide of any of SEQ ID NOS: 1, 9, 21, 24, 48, 54,

56, 57, 62, 63, 65, 76, 84 and 86, or a variant thereof. The at least one additional polypeptide is preferably a polypeptide of any of SEQ ID NOS: 63 and 65, or a variant thereof.

More specifically, in one embodiment, the invention therefore provides a composition comprising between four and thirteen polypeptides, consisting of:

a) at least one of the polypeptides of SEQ ID NOS. 83 and 82, or variants thereof, preferably SEQ ID NO. 83;

b) at least two of the polypeptides of SEQ ID NOS. 80, 81 and 85, or variants thereof; and optionally

c) at least one of the polypeptides of any of SEQ ID NOS: 5, 51, 52, 100, 101, 72, 73, and 74, or a variant thereof, preferably SEQ ID NOS: 51, 73, 100 and 104 or a variant thereof; and/or

d) at least one of the polypeptides of any of SEQ ID NOS: 1, 9, 21, 24, 48, 54, 56, 57, 62, 63, 65, 76, 84 and 86, or a variant thereof, preferably SEQ ID NOS: 63 and 65, or a variant thereof.

In other words, one specific embodiment of the invention provides a composition for use in the prevention or treatment of dust mite allergy by tolerisation comprising between four and thirteen peptide sequences, wherein the composition consists of:

a) at least one of the polypeptides with the following sequences:

HDM203B DLRQMRTVTPIRMQGGSGS (SEQ ID NO. 83) and

HDM203A DLRQMRTVTPIRMQGGCGS (SEQ ID NO. 82);

or a variant thereof, and;

b) at least two of the polypeptides with the following sequences:

HDM201 ESVKYVQSNGGAI (SEQ ID NO. 80);

HDM202 DEFKNRFLMSAEAFE (SEQ ID NO. 81); and

HDM205 SYYRYVAREQS (SEQ ID NO. 85)

or variants thereof and optionally;

c) at least one of the polypeptides with the following sequences:

HDM09A REALAQTHSAIAVI (SEQ ID NO. 5);

HDM03D RNQSLDLAEQELVDSASQH (SEQ ID NO. 51);

HDM03E RNQSLDLAEQELVDBASQH* (SEQ ID NO.52);

HDM03V EQELVDSASQHG (SEQ ID NO. 100);

HDM03W ELVDSASQHG (SEQ ID NO. 101);

HDM101 NYCQIYPPNVNKIREA (SEQ ID NO. 72);

HDM101A NYSQIYPPNVNKIREA (SEQ ID NO. 73); and

HDM101B NYBQIYPPNVNKIREA* (SEQ ID NO. 74)

or a variant thereof, and/or;

d) at least one of the polypeptides with the following sequences:

HDM01 IDLRQMRTVTPIR (SEQ ID NO. 1);

HDM21A KPFQLEAVFEANQNTK (SEQ ID NO. 9);

HDM48 TAIFQDTVRAEMTK (SEQ ID NO. 21);

HDM51A VDFKGELAMRNIEAR (SEQ ID NO. 24);

HDM01A IDLRQMRTVTPIRMQGGSG (SEQ ID NO. 48);

HDM06A RYVAREQSSRRP (SEQ ID NO. 54);

HDM07 PNVNKIREALAQT (SEQ ID NO. 56);

HDM19A DQVDVKDSANHEIKK (SEQ ID NO. 57);

HDM23C GLEVDVPGIDPNASH (SEQ ID NO. 62);

HDM26B GVLASAIATHAKIR (SEQ ID NO. 63);

HDM35A RGLKQMKRVGDANV (SEQ ID NO. 65);

HDM102A NAQRFGISNYSQI (SEQ ID NO. 76);

HDM204 SAYLAYRNQSLDLA (SEQ ID NO. 84); and

HDM206 DNGYGYFAANIDLMMIEE (SEQ ID NO. 86)

or a variant thereof. *B = aminobutyric acid.

It will be appreciated that (a) to (d) above represent stringent and highly selective criteria for the identification of suitable combinations of the invention. For example, if one were to select eight peptides at random from the sequences of the invention there would be nearly 100 billion possible combinations to choose from. By contrast, it is useful to consider an example of a combination of eight polypeptides in which the above criteria are applied. For example, consider a combination wherein the following polypeptides are selected:

i) any two of the polypeptides of SEQ ID NOS. 80, 81 and 85 and at least one

of the polypeptides of SEQ ID NOS. 82 and 83; and

ii) two further polypeptides selected from the polypeptides of any of SEQ ID NOS: 5, 51, 52, 72, 73, 74, 100 and 101; and finally

iii) two further polypeptides selected from the polypeptides of any of SEQ ID NOS: 1, 9, 21, 24, 48, 54, 56, 57, 62, 63, 65, 76, 84 and 86.

Based on such a selection, the number of possible combinations represents less than 0.0006% of the total available combinations if the criteria determined by the inventors are not applied.

On the basis of the above, a particularly preferred combination of the invention comprises or consists of the polypeptides of HDM201 (SEQ ID 80), HDM203B (SEQ ID 83), HDM205 (SEQ ID 85), HDM03W (SEQ ID 101), HDM101A (SEQ ID 73), HDM26B (SEQ ID 63), HDM35A (SEQ ID 65), and optionally SEQ ID NO. 24, or variants thereof.

Another preferred combination of the invention comprises or consists of the polypeptides of HDM201 (SEQ ID 80), HDM203B (SEQ ID 83), HDM205 (SEQ ID 85) and HDM03W (SEQ ID 101).

Subject to the above, the composition may optionally comprise further polypeptides up to a total of thirteen unique polypeptides. These further polypeptides relate to (i.e. are typically homologues and/or fragments of) the other sequences, i.e. SEQ ID NOS: 1 to 104, that are not amongst the polypeptides already selected. The further peptides are typically functional variants of one of the peptides of SEQ ID NO's 1 to 104. The further polypeptides may be identical to any of SEQ ID NOS: 1 to 104. The composition may therefore comprise up to thirteen different polypeptides as provided in any of SEQ ID NO: 1 to 104. However, the optional further polypeptides do not need to be 100% identical to any of SEQ ID NO: 1 to 104. They are preferably at least 65% identical to at least 9 (for example at least 10, 11, 12 or 13) or more contiguous amino acids in any of SEQ ID NO: 1 to 104, not already selected amongst the previously selected polypeptide(s). These contiguous amino acids may comprise a MHC class II epitope, for example which binds to any of the MHC molecules mentioned herein. In other words, the composition may optionally comprise further polypeptides up to a total of thirteen unique polypeptides,

wherein the further polypeptides:

(i)comprise a sequence having at least 65% sequence identity to at least 9 or more contiguous amino acids in any of SEQ ID NO: 1 to 104 above not selected in (a) to (d) above; and

(ii) are 9 to 30 amino acids in length.

wherein each different polypeptide is for simultaneous, separate or sequential use in the prevention or treatment of dust mite allergy by tolerisation.

1 to 104

In more detail therefore, the invention provides a product containing between four and thirteen polypeptides as defined in (a) to (d) above; and optionally:

(e) A polypeptide:

(i) comprising sequence having at least 65% sequence identity to at least 9 or more contiguous amino acids in any of SEQ ID NO: 1 to 104 not selected in a), to d) above; and

(ii) 9 to 30 amino acids in length; and optionally

(f) A polypeptide as defined in e), but additionally not selected in d); and optionally

(g)A polypeptide as defined in e), but additionally not selected in e) to f) above; and optionally

(h) A polypeptide as defined in e), but additionally not selected in e) to g) above; and optionally

(i) A polypeptide as defined in e), but additionally not selected in e) to h) above; and optionally

(j)A polypeptide as defined in e), but additionally not selected in e) to i) above; and optionally

(k)A polypeptide as defined in e), but additionally not selected in e) to j) above) above; and optionally

(l)A polypeptide as defined in e), but additionally not selected in e) to k) above; and optionally

(m) A polypeptide as defined in e), but additionally not selected in e) to l) above;

and optionally

(n) A polypeptide as defined in e), but additionally not selected in e) to m) above; and optionally

(o) A polypeptide as defined in e), but additionally not selected in e) to n) above; and optionally

(p) A polypeptide as defined in e), but additionally not selected in e) to o) above for simultaneous, separate or sequential use in the prevention or treatment of dust mite allergy by tolerisation.

Another embodiment of the invention is a composition for use in preventing or treating allergy to house dust mites by tolerisation comprising one or more polypeptide, wherein the polypeptide is selected from any of the following:

(i) a polypeptide of any of HDM203B (SEQ ID 83), HDM202 (SEQ ID 81), HDM201 (SEQ ID 80), HDM205 (SEQ ID 85), HDM203A (SEQ ID 82), SEQ ID NO's 1 to 79, 84, or 86 to 104 (that is any one of SEQ ID NO's. 1 to 104);or

(ii) a variant of a polypeptide according to (i), wherein said variant is a polypeptide of length 9 to 30 amino acids that comprises a region consisting of:

- any of the sequences of (i); or
- a sequence which has at least 65% homology to any of the sequences of (i) which sequence is capable of tolerising an individual to any of the sequences of (i), or

(iii) a variant of a polypeptide according to (i), wherein said variant is a polypeptide of length 9 to 30 amino acids that comprises a region consisting of a sequence that represents either:

- a fragment of any of the sequences of (i); or
- a homologue of a fragment of any of the sequences of (i),

which sequence is capable of tolerising an individual to any of the sequences of (i) and has a length of at least 9 amino acids, and wherein said homologue has at least 65% homology to any 9 contiguous amino acids in any of the sequences of (i).

The compositions or products of the invention may comprise variants of any of sequences defined above. The variant typically comprises 1, 2, 3 or more of the MHC class II epitopes present in the corresponding peptide of SEQ ID NO: 1 to 104.

Functional variants are mentioned herein. Such variants may be able to tolerise an individual to a class II MHC epitope present in the corresponding peptide of SEQ ID NO: 1 to 104, and thus it will typically comprise sequence that binds to the same MHC class II molecule and/or is recognised by a T cell which recognises the corresponding epitope in the polypeptide of SEQ ID NO: 1 to 104.

Variants of SEQ ID NO's 1 to 104 may be fragments derived by truncation. Truncation refers to the removal of one, two, three, four, five, six, seven, eight, nine, ten or more amino acids from the N and/or C-terminal ends of a polypeptide of SEQ ID NOS. 1 to 104. Examples of suitable truncations are provided for illustrative purposes in Example 5. In particular, truncations of SEQ ID NO. 81 are provided as SEQ ID NO's: 102 to 104. Similarly, a number of the preferred variants of HDM03 (SEQ ID NOS: 89 to 101) are truncations. Particularly preferred truncations of HDM03 are HDM03V and HDM 03W (SEQ ID 100 and 101).

Fragments may also be generated by one or more internal deletions, provided that the core 9 amino acids that makes up the T cell epitope is not substantially disrupted.

For example, a variant of SEQ ID NO: 1 may comprise a fragment of SEQ ID NO: 1, i.e. a shorter sequence. This may include a deletion of one, two, three, four, five, six, seven, eight, nine, ten or more amino acids from the N-terminal end of SEQ ID NO: 1 or from the C-terminal end of SEQ ID NO: 1. Such deletions may be made from both ends of SEQ ID NO: 1. A variant of SEQ ID NO: 1 may include additional amino acids (for example from the sequence of the parent protein from which the peptide derives) extending beyond the end(s) of SEQ ID NO: 1. A variant may include a combination of the deletions and additions discussed above. For example, amino acids may be deleted from one end of SEQ ID NO: 1, but additional amino acids from the full length parent protein sequence may be added at the other end of SEQ ID NO: 1. The same discussion of variants above also applies to SEQ ID NOS: 2 to 104.

A variant peptide may include one or more amino acid substitutions from the amino acid sequence of any of SEQ ID NOS: 1 to 104 or a fragment thereof. A variant peptide may comprise sequence having at least 65% sequence identity to at

least 9 or more contiguous amino acids in any of SEQ ID NOS: 1 to 104. More preferably a suitable variant may comprise at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% amino acid identity to at least 9 contiguous amino acids of any of SEQ ID NO: 1 to 104. This level of amino acid identity may be seen at any section of the peptide, although it is preferably the core region. The level of amino acid identity is over at least 9 contiguous amino acids but it may be at least 10, 11, 12, 13, 14, 15 or at least 16 or 17 amino acids, depending on the size of the peptides of comparison. Accordingly, any of the above-specified levels of identity may be across the entire length of sequence.

In connection with amino acid sequences, "sequence identity" refers to sequences which have the stated value when assessed using ClustalW (Thompson et al, Nucleic Acids Res. 1994 Nov 11;22(22):4673-80) with the following parameters: Pairwise alignment parameters -Method: accurate, Matrix: PAM, Gap open penalty: 10.00, Gap extension penalty: 0.10; Multiple alignment parameters -Matrix: PAM, Gap open penalty: 10.00, % identity for delay: 30, Penalize end gaps: on, Gap separation distance: 0, Negative matrix: no, Gap extension penalty: 0.20, Residue-specific gap penalties: on, Hydrophilic gap penalties: on, Hydrophilic residues: GPSNDQEKR. Sequence identity at a particular residue is intended to include identical residues which have simply been derivatised.

A variant peptide may comprise 1, 2, 3, 4, 5 or more, or up to 10 amino acid substitutions from any of SEQ ID NOS: 1 to 104. Substitution variants preferably involve the replacement of one or more amino acids with the same number of amino acids and making conservative amino acid substitutions. For example, an amino acid may be substituted with an alternative amino acid having similar properties, for example, another basic amino acid, another acidic amino acid, another neutral amino acid, another charged amino acid, another hydrophilic amino acid, another hydrophobic amino acid, another polar amino acid, another aromatic amino acid or another aliphatic amino acid. Some properties of the 20 main amino acids which can be used to select suitable substituents are as follows:

| | | | |
|---|---|---|---|
| Ala | aliphatic, hydrophobic, neutral | Met | hydrophobic, neutral |

| | | | |
|---|---|---|---|
| Cys | polar, hydrophobic, neutral | Asn | polar, hydrophilic, neutral |
| Asp | polar, hydrophilic, charged (-) | Pro | hydrophobic, neutral |
| Glu | polar, hydrophilic, charged (-) | Gln | polar, hydrophilic, neutral |
| Phe | aromatic, hydrophobic, neutral | Arg | polar, hydrophilic, charged (+) |
| Gly | aliphatic, neutral | Ser | polar, hydrophilic, neutral |
| His | aromatic, polar, hydrophilic, charged (+) | Thr | polar, hydrophilic, neutral |
| Ile | aliphatic, hydrophobic, neutral | Val | aliphatic, hydrophobic, neutral |
| Lys | polar, hydrophilic, charged(+) | Trp | aromatic, hydrophobic, neutral |
| Leu | aliphatic, hydrophobic, neutral | Tyr | aromatic, polar, hydrophobic |

Further variants include those in which instead of the naturally occurring amino acid the amino acid which appears in the sequence is a structural analog thereof. Amino acids used in the sequences may also be modified, e.g. labelled, providing the function of the peptide is not significantly adversely affected. Where the peptide has a sequence that varies from the sequence of any of SEQ ID NOS: 1 to 104 or a fragment thereof, the substitutions may occur across the full length of the sequence, within the sequence of any of SEQ ID NOS: 1 to 104 or outside the sequence of any of SEQ ID NOS: 1 to 104. For example, the variations described herein, such as additions, deletions, substitutions and modifications, may occur within the sequence of any of SEQ ID NOS: 1 to 104. A variant peptide may comprise or consist essentially of the amino acid sequence of any of SEQ ID NOS: 1 to 104 in which one, two, three, four or more amino acid substitutions have been made. A variant peptide may comprise a fragment of the parent protein that is larger than any of SEQ ID NOS: 1 to 104. In this embodiment, the variations described herein, such as substitutions and modifications, may occur within and/or outside the sequence of any of SEQ ID NOS: 1 to 104.

The variant peptides of the invention are 9 to 30 amino acids in length inclusive. Preferably, they may be from 9 to 20 or more preferably 13 to 17 amino acids in length. The peptides may be the same length as the peptide sequences in any one of SEQ ID NOS: 1 to 20.

The peptides may be chemically derived from the polypeptide allergen, for example by proteolytic cleavage or can be derived in an intellectual sense from the polypeptide allergen, for example by making use of the amino acid sequence of the polypeptide allergen and synthesising peptides based on the sequence. Peptides may be synthesised using methods well known in the art.

Where polypeptides comprise residues which are typically difficult to preserve during manufacture, these residues may be replaced. For example, glutamate spontaneously forms pyroglutamate in solution particularly when present at the N terminus of a peptide. Thus, residues of the peptides of the invention which correspond to glutamate in the sequence of a native allergen protein sequence may be replaced with pyrogluatmate in the peptides of the invention when such residues are present at the N terminus of a peptide.

The term "peptide" includes not only molecules in which amino acid residues are joined by peptide (-CO-NH-) linkages but also molecules in which the peptide bond is reversed. Such retro-inverso peptidomimetics may be made using methods known in the art, for example such as those described in Meziere *et al* (1997) J. Immunol.159, 3230-3237. This approach involves making pseudopeptides containing changes involving the backbone, and not the orientation of side chains. Meziere *et al* (1997) show that, at least for MHC class II and T helper cell responses, these pseudopeptides are useful. Retro-inverse peptides, which contain NH-CO bonds instead of CO-NH peptide bonds, are much more resistant to proteolysis.

Similarly, the peptide bond may be dispensed with altogether provided that an appropriate linker moiety which retains the spacing between the carbon atoms of the amino acid residues is used; it is particularly preferred if the linker moiety has substantially the same charge distribution and substantially the same planarity as a peptide bond. It will also be appreciated that the peptide may conveniently be blocked at its N-or C-terminus so as to help reduce susceptibility to exoproteolytic digestion. For example, the N-terminal amino group of the peptides may be protected by reacting with a carboxylic acid and the C-terminal carboxyl group of the peptide may be protected by reacting with an amine. Other examples of modifications include glycosylation and phosphorylation. Another potential

modification is that hydrogens on the side chain amines of R or K may be replaced with methylene groups ($-NH_2 \rightarrow -NH(Me)$ or $-N(Me)_2$).

Analogues of peptides according to the invention may also include peptide variants that increase or decrease the peptide's half-life *in vivo*. Examples of analogues capable of increasing the half-life of peptides used according to the invention include peptoid analogues of the peptides, D-amino acid derivatives of the peptides, and peptide-peptoid hybrids. A further embodiment of the variant polypeptides used according to the invention comprises D-amino acid forms of the polypeptide. The preparation of polypeptides using D-amino acids rather than L-amino acids greatly decreases any unwanted breakdown of such an agent by normal metabolic processes, decreasing the amounts of agent which needs to be administered, along with the frequency of its administration.

The peptides provided by the present invention may be derived from splice variants of the parent proteins encoded by mRNA generated by alternative splicing of the primary transcripts encoding the parent protein chains. The peptides may also be derived from amino acid mutants, glycosylation variants and other covalent derivatives of the parent proteins which retain at least an MHC-binding property of the allergens. Exemplary derivatives include molecules wherein the peptides of the invention are covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid. Further included are naturally occurring variants of the parent proteins found in different mites. Such a variant may be encoded by an allelic variant or represent an alternative splicing variant.

Variants as described above may be prepared during synthesis of the peptide or by post- production modification, or when the peptide is in recombinant form using the known techniques of site- directed mutagenesis, random mutagenesis, or enzymatic cleavage and/or ligation of nucleic acids.

In accordance with the invention, the further peptides that the composition may comprise are preferably functional variants of any of SEQ ID NOS: 1 to 104. That is, the peptides are preferably capable of inducing an immune response. In particular, the peptides are preferably capable of inducing cytokine production in

house dust mite allergic individuals. Typically, the composition of the invention will therefore comprise at least one polypeptide or variant thereof which produces a cytokine response in greater than 30, 35, 40%, preferably 45% or 50% of individuals in a population of house dust mite allergic individuals. The number of individuals in a panel of dust mite allergic individuals may be any number greater than one, for example at least 20, 30, 40, 50, 80, or at least 100 individuals. Preferably the composition comprises at least two, three or most preferably four such peptides. Preferably the cytokine response is production of IL13 or IFN-gamma. Cytokine production may be measured by any suitable method. Production of a cytokine is typically considered to have occurred in response to a peptide if the level of cytokine produced in the presence of the peptide is at least 2, 3, 4 or 5 fold above the background level of said cytokine that is produced in the absence of a stimulus (i.e. the level produced by the same individual in the absence of the peptide or any other stimulus). Alternatively, production of a cytokine may be considered to have occurred if the amount of cytokine produced exceeds a recognised limit, typically 90, 95, or preferably 100 pg/ml, typically from a sample of approximately $1.25x10^6$ cells in 250µl.

Suitable methods for measuring cytokine production typically include measuring the cytokine release from peripheral blood mononuclear cells (PBMCs) from a taken sample from a subject. The sample is typically blood or serum. Cytokine release from PBMCs is measured after incubating the cells in the presence of a given peptide. Supernatants from the incubation mixture are then tested for the presence of a cytokine, using any suitable assay, for example an ELISA, ELISPOT assay or flow cytometric assay. Particularly preferred methods include Multiplex bead array assays as described in, for example de Jager *et al;* Clinical and Diagnostic Laboratory Immunology, 2003, Vol 10(1) p. 133-139.Typically, the composition may comprise at least one additional peptide or variant thereof that is not amongst the polypeptides already selected, upto a total of thirteen different peptides, which produces a cytokine response in greater than 20%, 25%, preferably 30%, 35% or 40% of individuals in a population of house dust mite allergic individuals.

The composition may further comprise one or more additional peptides or

variants thereof that are not amongst the polypeptides already selected, upto a total of thirteen different peptides, which produce a cytokine response in greater than 10%, 15%, preferably 20% of individuals in a population of house dust mite allergic individuals.

Suitable variants capable of binding to TCRs may be derived empirically or selected according to known criteria. Within a single peptide there are certain residues which contribute to binding within the MHC antigen binding groove and other residues which interact with hypervariable regions of the T cell receptor (Allen et al (1987) Nature 327: 713-5).

Within the residues contributing to T cell receptor interaction, a hierarchy has been demonstrated which pertains to dependency of T cell activation upon substitution of a given peptide residue. Using peptides which have had one or more T cell receptor contact residues substituted with a different amino acid, several groups have demonstrated profound effects upon the process of T cell activation. Evavold & Allen (1991) Nature 252: 1308-10) demonstrated the dissociation of T cell proliferation and cytokine production. In this *in vitro* model, a T cell clone specific for residues 64-76 of haemoglobin (in the context of I-$E^k$), was challenged with a peptide analogue in which a conservative substitution of aspartic acid for glutamic acid had been made. This substitution did not significantly interfere with the capacity of the analogue to bind to I-$E^k$.

Following *in vitro* challenge of a T cell clone with this analogue, no proliferation was detected although IL-4 secretion was maintained, as was the capacity of the clone to help B cell responses. In a subsequent study the same group demonstrated the separation of T cell-mediated cytolysis from cytokine production. In this instance, the former remained unaltered while the latter was impaired. The efficacy of altered peptide ligands *in vivo* was initially demonstrated in a murine model of EAE (experimental allergic encephalomyelitis) by McDevitt and colleagues (Smilek *et al* (1991) Proc Natl Acad Sci USA 88 : 9633-9637). In this model EAE is induced by immunisation with the encephalitogenic peptide Ac1-11 of MBP (myelin basic protein). Substitution at position four (lysine) with an alanine residue generated a peptide which bound well to its restricting element ($A\alpha^u A\beta^u$), but which was non-

immunogenic in the susceptible PL/JxSJLF1 strain and which, furthermore prevented the onset of EAE when administered either before or after immunisation with the encephalitogenic peptide. Thus, residues can be identified in peptides which affect the ability of the peptides to induce various functions of T-cells.

Advantageously, peptides may be designed to favour T-cell proliferation and induction of desensitisation. Metzler and Wraith have demonstrated improved tolerogenic capacity of peptides in which substitutions increasing peptide-MHC affinity have been made (Metzler & Wraith(1993) Int Immunol ~ : 1159-65). That an altered peptide ligand can cause long-term and profound anergy in cloned T cells was demonstrated by Sloan-Lancaster *et al* (1993) Nature 363: 156-9.

The compositions of the invention are capable of inducing a late phase response in an individual that is sensitised to the allergens. The term "late phase response" includes the meaning as set forth in Allergy and Allergic Diseases (1997) A. B. Kay (Ed.), Blackwell Science, pp 1113-1130. The late phase response may be any late phase response (LPR). Preferably, the peptides are capable of inducing a late asthmatic response (LAR) or a late rhinitic response, or a late phase skin response or a late phase ocular response. Whether or not a particular peptide can give rise to a LPR can be determined using methods well known in the art; a particularly preferred method is that described in Cromwell O, Durham SR, Shaw RJ, Mackay J and Kay AB. Provocation tests and measurements of mediators from mast cells and basophils in asthma and allergic rhinitis. In: Handbook of Experimental Immunology (4) Chapter 127, Editor: Weir DM, Blackwell Scientific Publications, 1986.

Thus, preferably, the individual peptides of the invention are able to induce a LPR in an individual who has been sensitised to the allergens. Whether or not an individual has been sensitised to the allergens may be determined by well known procedures such as skin prick testing with solutions of allergen extracts, induction of cutaneous LPRs, clinical history, allergen challenge and radioallergosorbent test (RAST) for measurement of allergen specific IgE. Whether or not a particular individual is expected to benefit from treatment may be determined by the physician based, for example, on such tests.

Desensitising or tolerising an individual to the allergens means inhibition or

dampening of allergic tissue reactions induced by the allergens in appropriately sensitised individuals. It has been shown that T cells can be selectively activated, and then rendered unresponsive. Moreover the anergising or elimination of these T-cells leads to desensitisation of the patient for a particular allergen. The desensitisation manifests itself as a reduction in response to an allergen or allergen-derived peptide, or preferably an elimination of such a response, on second and further administrations of the allergen or allergen-derived peptide. The second administration may be made after a suitable period of time has elapsed to allow desensitisation to occur; this is preferably any period between one day and several weeks. An interval of around two weeks is preferred.

Although the compositions of the invention are able to induce a LPR in a dust mite allergic individual, it should be appreciated that when a composition is used to treat a patient it is preferable that a sufficiently low concentration of the composition is used such that no observable LPR will occur but the response will be sufficient to partially desensitise the T cells such that the next (preferably higher) dose may be given, and so on. In this way the dose is built up to give full desensitisation but often without ever inducing a LPR in the patient. Although, the composition or peptide is able to do so at a higher concentration than is administered.

The compositions of the invention preferably are capable of inducing a late phase response in 50 % or more of a panel of dust mite allergic individuals from the population. More preferably, the compositions are capable of inducing a LPR in 55% or more, 60 % or more, 65 % or more, 70% or more, 75% or more, 80% or more, 85% or more, or 90 % or more of sensitized individuals in a panel. Whether or not the compositions are able to induce a LPR in a certain percentage of a panel of subjects can be determined by methods which are well known in the art.

It will be understood that the peptides of the invention comprise a T cell epitope that consists of a core 9 amino acids which are the minimal essential sequence required for MHC class II binding. However, the peptides may also comprise additional residues flanking the core 9 amino acids. The peptides may therefore comprise a region containing a T cell epitope, in which some residues may be modified without affecting the function of the epitope. Accordingly, functional

variants of the peptides as defined above include peptides which are altered to improve their solubility relative to the native sequence of the peptides. Improved solubility is advantageous for the tolerisation of subjects to allergens from which the peptides of the invention derive, since administration of poorly soluble agents to subjects causes undesirable, non-tolerising inflammatory responses. The solubility of the peptides may be improved by altering the residues which flank the region containing a T cell epitope. A peptide of the invention may be engineered to be more soluble such that it comprises:

i) N terminal to the residues of the peptide which flank a T cell epitope: one to six contiguous amino acids corresponding to the two to six contiguous amino acids immediately N terminal to said residues in the sequence of the protein from which the peptide derives; and/or

ii) C terminal to the residues of the peptide which flank a T cell epitope: one to six contiguous amino acids corresponding to the one to six contiguous amino acids immediately C terminal to the said residues in the sequence of the protein from which the peptide derives; or

iii)both N and C terminal to the residues of the peptide which flank a T cell epitope, at least one amino acid selected from arginine, lysine, histidine, glutamate and aspartate.

Optionally, the peptides may additionally be engineered to be more soluble such that:

i) any cysteine residues in the native sequence of the peptide are replaced with serine or 2-aminobutyric acid; and /or

ii) any residues at the N or C terminus of the native sequence of the peptide, which are not comprised in a T cell epitope, are deleted; and/or

iii) any two consecutive amino acids comprising the sequence Asp-Gly in the upto four amino acids at the N or C terminus of the native sequence of the peptide, which are not comprised in a T cell epitope, are deleted.

*Nucleic acids and vectors*

The individual peptides that make up the compositions and products of the

invention may be administered directly, or may be administered indirectly by expression from an encoding sequence. For example, a polynucleotide may be provided that encodes a peptide of the invention, such as any of the peptides described above. A peptide of the invention may thus be produced from or delivered in the form of a polynucleotide which encodes, and is capable of expressing, it. Any reference herein to the use, delivery or administration of a peptide of the invention is intended to include the indirect use, delivery or administration of such a peptide via expression from a polynucleotide that encodes it.

The terms "nucleic acid molecule" and "polynucleotide" are used interchangeably herein and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Non-limiting examples of polynucleotides include a gene, a gene fragment, messenger RNA (mRNA), cDNA, recombinant polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide of the invention may be provided in isolated or purified form.
A nucleic acid sequence which "encodes" a selected polypeptide is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. For the purposes of the invention, such nucleic acid sequences can include, but are not limited to, cDNA from viral, prokaryotic or eukaryotic mRNA, genomic sequences from viral or prokaryotic DNA or RNA, and even synthetic DNA sequences. A transcription termination sequence may be located 3' to the coding sequence.

Polynucleotides of the invention can be synthesised according to methods well known in the art, as described by way of example in Sambrook et al (19104, Molecular Cloning - a laboratory manual; Cold Spring Harbor Press).

The polynucleotide molecules of the present invention may be provided in the form of an expression cassette which includes control sequences operably linked to the inserted sequence, thus allowing for expression of the peptide of the invention *in*

*vivo* in a targeted subject. These expression cassettes, in turn, are typically provided within vectors (e.g., plasmids or recombinant viral vectors) which are suitable for use as reagents for nucleic acid immunization. Such an expression cassette may be administered directly to a host subject. Alternatively, a vector comprising a polynucleotide of the invention may be administered to a host subject. Preferably the polynucleotide is prepared and/or administered using a genetic vector. A suitable vector may be any vector which is capable of carrying a sufficient amount of genetic information, and allowing expression of a peptide of the invention.

The present invention thus includes expression vectors that comprise such polynucleotide sequences. Thus, the present invention provides a vector for use in preventing or treating allergy to dust mites by tolerisation comprising four or more polynucleotide sequences which encode different polypeptides of the invention and optionally one or more further polynucleotide sequences which encode different polypeptides as defined herein. The vector may comprise 4, 5, 6 or 7 polynucleotide sequences which encode different polypeptides of the invention.

Furthermore, it will be appreciated that the compositions and products of the invention may comprise a mixture of polypeptides and polynucleotides. Accordingly, the invention provides a composition or product as defined herein, wherein in place of any one of the polypeptide is a polynucleotide capable of expressing said polypeptide.

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

Thus, a polypeptide of the invention may be provided by delivering such a vector to a cell and allowing transcription from the vector to occur. Preferably, a polynucleotide of the invention or for use in the invention in a vector is operably linked to a control sequence which is capable of providing for the expression of the

coding sequence by the host cell, i.e. the vector is an expression vector.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given regulatory sequence, such as a promoter, operably linked to a nucleic acid sequence is capable of effecting the expression of that sequence when the proper enzymes are present. The promoter need not be contiguous with the sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the nucleic acid sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

A number of expression systems have been described in the art, each of which typically consists of a vector containing a gene or nucleotide sequence of interest operably linked to expression control sequences. These control sequences include transcriptional promoter sequences and transcriptional start and termination sequences. The vectors of the invention may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. A "plasmid" is a vector in the form of an extrachromosomal genetic element. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a resistance gene for a fungal vector. Vectors may be used *in vitro*, for example for the production of DNA or RNA or used to transfect or transform a host cell, for example, a mammalian host cell. The vectors may also be adapted to be used *in vivo*, for example to allow *in vivo* expression of the polypeptide.

A "promoter" is a nucleotide sequence which initiates and regulates transcription of a polypeptide-encoding polynucleotide. Promoters can include inducible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein, etc.), repressible promoters (where expression of a polynucleotide sequence operably linked to the promoter is repressed by an analyte, cofactor, regulatory protein, etc.), and constitutive promoters. It is intended that the term "promoter" or "control

element" includes full-length promoter regions and functional (e.g., controls transcription or translation) segments of these regions.

A polynucleotide, expression cassette or vector according to the present invention may additionally comprise a signal peptide sequence. The signal peptide sequence is generally inserted in operable linkage with the promoter such that the signal peptide is expressed and facilitates secretion of a polypeptide encoded by coding sequence also in operable linkage with the promoter.

Typically a signal peptide sequence encodes a peptide of 10 to 30 amino acids for example 15 to 20 amino acids. Often the amino acids are predominantly hydrophobic. In a typical situation, a signal peptide targets a growing polypeptide chain bearing the signal peptide to the endoplasmic reticulum of the expressing cell. The signal peptide is cleaved off in the endoplasmic reticulum, allowing for secretion of the polypeptide via the Golgi apparatus. Thus, a peptide of the invention may be provided to an individual by expression from cells within the individual, and secretion from those cells.

Alternatively, polynucleotides of the invention may be expressed in a suitable manner to allow presentation of a peptide of the invention by an MHC class II molecule at the surface of an antigen presenting cell. For example, a polynucleotide, expression cassette or vector of the invention may be targeted to antigen presenting cells, or the expression of encoded peptide may be preferentially stimulated or induced in such cells.

Polynucleotides of interest may be used *in vitro, ex vivo* or *in vivo* in the production of a peptide of the invention. Such polynucleotides may be administered or used in the prevention or treatment of allergy by tolerisation.

Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859 and 5,5104,466. The nucleic acid molecule can be introduced directly into the recipient subject, such as by standard intramuscular or intradermal injection; transdermal particle delivery; inhalation; topically, or by oral, intranasal or mucosal modes of administration. The molecule alternatively can be introduced *ex vivo* into cells that have been removed from a subject. For example, a polynucleotide, expression cassette or vector of the invention may be introduced into

APCs of an individual *ex vivo*. Cells containing the nucleic acid molecule of interest are re-introduced into the subject such that an immune response can be mounted against the peptide encoded by the nucleic acid molecule. The nucleic acid molecules used in such immunization are generally referred to herein as "nucleic acid vaccines."

The polypeptides, polynucleotides, vectors or cells of the invention may be present in a substantially isolated form. They may be mixed with carriers or diluents which will not interfere with their intended use and still be regarded as substantially isolated. They may also be in a substantially purified form, in which case they will generally comprise at least 90%, e.g. at least 95%, 98% or 99%, of the proteins, polynucleotides, cells or dry mass of the preparation.

*Antigen presenting cells (APCs)*

The invention encompasses the use *in vitro* of a method of producing a population of APCs that present the peptides of the invention on their surface, that may be subsequently used in therapy. Such a method may be carried out *ex vivo* on a sample of cells that have been obtained from a patient. The APCs produced in this way therefore form a pharmaceutical agent that can be used in the treatment or prevention of dust mite allergy by tolerisation. The cells should be accepted by the immune system of the individual because they derive from that individual. Delivery of cells that have been produced in this way to the individual from whom they were originally obtained, thus forms a therapeutic embodiment of the invention.

*Formulations and compositions*

The peptides, polynucleotides, vectors and cells of the invention may be provided to an individual either singly or in combination. Each molecule or cell of the invention may be provided to an individual in an isolated, substantially isolated, purified or substantially purified form. For example, a peptide of the invention may be provided to an individual substantially free from the other peptides.

Whilst it may be possible for the peptides, polynucleotides or compositions according to the invention to be presented in raw form, it is preferable to present

them as a pharmaceutical formulation. Thus, according to a further aspect of the invention, the present invention provides a pharmaceutical formulation for use in preventing or treating allergy to dust mites by tolerisation comprising a composition, vector or product according to the invention together with one or more pharmaceutically acceptable carriers or diluents and optionally one or more other therapeutic ingredients. The carrier (s) must be 'acceptable' in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof. Typically, carriers for injection, and the final formulation, are sterile and pyrogen free.

Formulation of a composition comprising the peptide, polynucleotides or cells of the invention can be carried out using standard pharmaceutical formulation chemistries and methodologies all of which are readily available to the reasonably skilled artisan.

For example, compositions containing one or more molecules or cells of the invention can be combined with one or more pharmaceutically acceptable excipients or vehicles. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances and the like, may be present in the excipient or vehicle. These excipients, vehicles and auxiliary substances are generally pharmaceutical agents that do not induce an immune response in the individual receiving the composition, and which may be administered without undue toxicity. Pharmaceutically acceptable excipients include, but are not limited to, liquids such as water, saline, polyethyleneglycol, hyaluronic acid and ethanol. Pharmaceutically acceptable salts can also be included therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients, vehicles and auxiliary substances is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Such compositions may be prepared, packaged, or sold in a form suitable for bolus administration or for continuous administration. Injectable compositions may be prepared, packaged, or sold in unit dosage form, such as in ampoules or in multi-dose containers containing a preservative. Compositions include, but are not limited

to, suspensions, solutions, emulsions in oily or aqueous vehicles, pastes, and implantable sustained-release or biodegradable formulations. Such compositions may further comprise one or more additional ingredients including, but not limited to, suspending, stabilizing, or dispersing agents. In one embodiment of a composition for parenteral administration, the active ingredient is provided in dry (for e.g., a powder or granules) form for reconstitution with a suitable vehicle (e. g., sterile pyrogen-free water) prior to parenteral administration of the reconstituted composition. The pharmaceutical compositions may be prepared, packaged, or sold in the form of a sterile injectable aqueous or oily suspension or solution. This suspension or solution may be formulated according to the known art, and may comprise, in addition to the active ingredient, additional ingredients such as the dispersing agents, wetting agents, or suspending agents described herein. Such sterile injectable formulations may be prepared using a non-toxic parenterally-acceptable diluent or solvent, such as water or 1,3-butane diol, for example. Other acceptable diluents and solvents include, but are not limited to, Ringer's solution, isotonic sodium chloride solution, and fixed oils such as synthetic mono-or di-glycerides. Other parentally-administrable compositions which are useful include those which comprise the active ingredient in microcrystalline form, in a liposomal preparation, or as a component of a biodegradable polymer systems. Compositions for sustained release or implantation may comprise pharmaceutically acceptable polymeric or hydrophobic materials such as an emulsion, an ion exchange resin, a sparingly soluble polymer, or a sparingly soluble salt.

Alternatively, the peptides or polynucleotides of the present invention may be encapsulated, adsorbed to, or associated with, particulate carriers. Suitable particulate carriers include those derived from polymethyl methacrylate polymers, as well as PLG microparticles derived from poly(lactides) and poly(lactide-co-glycolides). See, e.g., Jeffery et al. (1993) Pharm. Res. 10:362-368. Other particulate systems and polymers can also be used, for example, polymers such as polylysine, polyarginine, polyornithine, spermine, spermidine, as well as conjugates of these molecules.

The formulation of any of the peptides, polynucleotides or cells mentioned

herein will depend upon factors such as the nature of the substance and the method of delivery. Any such substance may be administered in a variety of dosage forms. It may be administered orally (e.g. as tablets, troches, lozenges, aqueous or oily suspensions, dispersible powders or granules), parenterally, epicutaneously, subcutaneously, by inhalation, intravenously, intramuscularly, intrasternally, transdermally, intradermally, sublingually, instranasally, buccally or by infusion techniques. The substance may also be administered as suppositories. A physician will be able to determine the required route of administration for each particular individual.

The compositions of formulations of the invention will comprise a suitable concentration of each peptide/polynucleotide/cell to be effective without causing adverse reaction. Typically, the concentration of each peptide in the composition will be in the range of 0.03 to 200 nmol/ml. More preferably in the range of 0.3 to 200 nmol/ml, 3 to 180 nmol/ml, 10 to 150 nmol/ml or 30 to 120 nmol/ml. The composition or formulations should have a purity of greater than 95% or 98% or a purity of at least 99%.

In one embodiment, therefore, the peptides, polynucleotides, cells or compositions of the invention are used for therapy in combination with one or more other therapeutic agents. The agents may be administered separately, simultaneously or sequentially. They may be administered in the same or different compositions. Accordingly, in a method of the invention, the subject may also be treated with a further therapeutic agent.

A composition may therefore be formulated which comprises a molecule and/or cell of the invention and also one or more other therapeutic molecules. A composition of the invention may alternatively be used simultaneously, sequentially or separately with one or more other therapeutic compositions as part of a combined treatment.

*Therapeutic methods and individual to be treated*

The present invention relates to peptides, polynucleotides, vectors and cells that are capable of desensitising or tolerising human individuals to the allergens

described above and are therefore useful in the prevention or treatment of dust mite allergy. The invention provides compositions, products, vectors and formulations for use in preventing or treating allergy to dust mites by tolerisation. The invention also provides a method of tolerising or desensitizing a dust mite allergic individual comprising administering, either singly or in combination the polypeptides/polynucleotides/cells of the invention as described above.

The individual to be treated or provided with the composition or formulation of the invention is preferably human. It will be appreciated that the individual to be treated may be known to be sensitised to the allergens, at risk of being sensitised or suspected of being sensitised. The individual can be tested for sensitisation using techniques well known in the art and as described herein. Alternatively, the individual may have a family history of allergy to dust mites. It may not be necessary to test an individual for sensitisation to dust mites because the individual may display symptoms of allergy when exposed to dust mites. By exposure is meant proximity to, for example, an item of clothing, a mattress, pillow, pillow case, sheet, blanket or other bedding material which has not been washed at greater than 50°C for more than approximately one week, or a carpet, curtain or upholstered item of furniture which has not been vacuum cleaned for more than approximately one week. By proximity is meant 10 metres or less, 5 metres or less, 2 metres or less, 1 metre or less, or 0 metres from the items described above. Symptoms of allergy can include itchy eyes, runny nose, breathing difficulties, red itchy skin or rash.

The individual to be treated may be of any age. However, preferably, the individual may be in the age group of 1 to 90, 5 to 60, 10 to 40, or more preferably 18 to 35.

Preferably, the individual to be treated is from a population that has MHC allele frequencies within the range of frequencies that are representative of the Caucasian population. Reference population allele frequencies for 11 common DRB1 allele families are shown in Table 1 (Data from HLA Facts Book, Parham and Barber).

Table 1

| DRB1 | 1 | 3 | 4 | 7 | 8 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| % | 6.4 | 14.7 | 15.7 | 8.8 | 3.4 | 8.3 | 3.9 | 14.7 | 2.9 | 17.6 | 2.5 |
| Reference population % | 9.4 | 11.1 | 12.8 | 13.2 | 3.7 | 13.4 | 2.3 | 10.2 | 3.2 | 10.7 | 3.6 |

Reference frequencies were obtained by analysis of multiple studies reporting frequencies and the figures shown are mean values. Preferably therefore, the individual to be treated is from a population that has equivalent MHC allele frequencies as the reference population for the alleles referred to Table 1 (such as for at least 1, 2, 3, 4, 5 or all of the alleles), for example within the ranges of those figures plus or minus 1, 2, 3, 5, 10, 15 or 20%.

Preferably the individual is from a population where the allele frequencies of the following DRB1 alleles is:

4 – at least 9%

7 – at least 10%

11 – at least 8%.

The individual may have had allergy to dust mites for at least 2 weeks, 1 month, 6 months, 1 year or 5 years. The individual may suffer from a rash, nasal congestion, nasal discharge and/or coughing caused by the allergy. The individual may or may not have been administered with other compositions/compounds which treat dust mite allergy. The individual may live in a geographical region which experiences a daily average relative humidity greater than 50%, preferably 55%, 60%, 65%, 70%, 75%, 80% or 90%. The individual may live in a geographical region known to support dust mite populations, for example the eastern half of the United States (and major western coastal cities of the United States), populous areas of Canada, western Europe, Japan, Korea, and coastal areas of South America, Australia and South Africa.

*Combination immunotherapy*

Since many individuals are allergic, or may require desensitizing to several

polypeptide antigens, the current invention also provides means of desensitizing individuals that are allergic to multiple antigens. "Tolerance" induced in an individual to a first polypeptide antigen or allergen can create in the individual a "tolergeneic environment" wherein inappropriate immune responses to other antigens can be downregulated in order to provide tolerance to other antigens.

This finding means that individuals allergic to multiple allergens can be treated in a greatly reduced time period, and that individuals seriously allergic to some allergens (e.g., peanuts) but more mildly allergic to other allergens (e.g., cat dander) can benefit from a therapy wherein tolerance to the milder allergen is established and then this tolergeneic environment is used to provide tolerance to the other, more extreme allergen. In addition, individuals suffering from an autoimmune disorder who are additionally sensitised (or otherwise immune) to an unrelated antigen or allergen can benefit from a treatment regime wherein tolerance to the unrelated antigen or allergen is first established and then this tolergeneic environment is used to provide tolerance to the autoantigen associated with the autoimmune disorder.

A method is therefore provided for desensitising a dust mite allergic individual to dust mite allergen as described above and one or more further different polypeptide antigens. The method entails, in a first step, administering to the individual a composition/product/formulation (primary composition) according to the invention as described herein and wherein the administration is carried out in a manner sufficient to generate a hyporesponsive state against dust mite allergen. Once a hyporesponsive state has been established toward dust mite allergen, or at least a shift toward desensitisation has occurred, the method entails administration of a secondary composition comprising a second, different polypeptide antigen to which the individual is to be sensitised. Administration of the secondary composition is carried out in such a way as to take advantage of the tolergeneic environment established by use of the primary composition, where it is now possible to establish tolerance to the second, different polypeptide antigen. The secondary composition is coadministered with either the first primary composition or a larger fragment of Feld1. By "coadministered" it is meant either the simultaneous or concurrent

administration, e.g., when the two are present in the same composition or administered in separate compositions at nearly the same time but at different sites, as well as the delivery of polypeptide antigens in separate compositions at different times. For example, the secondary composition may be delivered prior to or subsequent to delivery of the first composition at the same or a different site. The timing between deliveries can range from about several seconds apart to about several minutes apart, several hours apart, or even several days apart. Furthermore, different delivery methods can be employed.

The second polypeptide antigen is preferably an allergen different to the dust mite allergen. Suitable allergens for use in the methods of the invention can of course be obtained and/or produced using known methods. Classes of suitable allergens include, but are not limited to, other dust mite allergens, pollens, animal dander (especially cat dander), grasses, molds, dusts, antibiotics, stinging insect venoms, and a variety of environmental (including chemicals and metals), drug and food allergens. Common tree allergens include pollens from cottonwood, popular, ash, birch, maple, oak, elm, hickory, and pecan trees; common plant allergens include those from mugwort, ragweed, English plantain, sorrel-dock and pigweed; plant contact allergens include those from poison oak, poison ivy and nettles; common grass allergens include rye grass, Timothy, Johnson, Bermuda, fescue and bluegrass allergens; common allergens can also be obtained from molds or fungi such as Alternaria, Fusarium, Hormodendrum, Aspergillus, Micropolyspora, Mucor and thermophilic actinomycetes; epidermal allergens can be obtained from house or organic dusts (typically fungal in origin), or from animal sources such as feathers, and dog dander; common food allergens include milk and cheese (diary), egg, wheat, nut (e.g., peanut), seafood (e.g., shellfish), pea, bean and gluten allergens; common environmental allergens include metals (nickel and gold), chemicals (formaldehyde, trinitrophenol and turpentine), Latex, rubber, fiber (cotton or wool), burlap, hair dye, cosmetic, detergent and perfume allergens; common drug allergens include local anesthetic and salicylate allergens; antibiotic allergens include penicillin, tetracycline and sulfonamide allergens; and common insect allergens include bee, wasp and ant venom, and cockroach calyx allergens. Particularly well characterized allergens

include, but are not limited to, the major cat allergen Fel d1, bee venom phospholipase A2 (PLA) (Akdis et al. (1996) *J. Clin. Invest.* 98:1676-1683), birch pollen allergen Bet v 1 (Bauer et al. (1997) *Clin. Exp. Immunol.* 107:536-541), and the multi-epitopic recombinant grass allergen rKBG8.3 (Cao et al. (1997) *Immunology* 90:46-51). These and other suitable allergens are commercially available and/or can be readily prepared as extracts following known techniques.

Preferably, the second polypeptide allergen is selected from the list of allergen sequences and database accession numbers (NCBI Entrez accession numbers) below. NCBI is the National Center for Biotechnology information and is a division of the US National Institutes of Health. The NCBI web site, from which access to the database may be sought, is www.ncbi.nlm.nih.gov/. Allergen sequences and database accession numbers (NCBI Entrez accession numbers):

House dust mite

Dermatophagoides pteronyssinus

Der p 1

MKIVLAIASLLALSAVYARPSSIKTFEEYKKAFNKSYATFEDEEAARKNFLES
VKYVQSNGGAINHLSDLSLDEFKNRFLMSAEAFEHLKTQFDLNAETNACSIN
GNAPAEIDLRQMRTVTPIRMQGGCGSCWAFSGVAATESAYLAYRNQSLDLA
EQELVDCASQHGCHGDTIPRGIEYIQHNGVVQESYYRYVAREQSCRRPNAQR
FGISNYCQIYPPNVNKIREALAQTHSAIAVIIGIKDLDAFRHYDGRTIIQRDNGY
QPNYHAVNIVGYSNAQGVDYWIVRNSWDTNWGDNGYGYFAANIDLMMIEE
YPYVVIL

Der p 2

MMYKILCLSLLVAAVARDQVDVKDCANHEIKKVLVPGCHGSEPCIIHRGKPF
QLEAVFEANQNTKTAKIEIKASIDGLEVDVPGIDPNACHYMKCPLVKGQQYD
IKYTWNVPKIAPKSENVVVTVKVMGDDGVLACAIATHAKIRD

Der p 3

MIIYNILIVLLLAINTLANPILPASPNATIVGGEKALAGECPYQISLQSSSHFCGG
TILDEYWILTAAHCVAGQTASKLSIRYNSLKHSLGGEKISVAKIFAHEKYDSY
QIDNDIALIKLKSPMKLNQKNAKAVGLPAKGSDVKVGDQVRVSGWGYLEEG
SYSLPSELRRVDIAVVSRKECNELYSKANAEVTDNMICGGDVANGGKDSCQ
GDSGGPVVDVKNNQVVGIVSWGYGCARKGYPGVYTRVGNFIDWIESKRSQ

Der p 4
KYXNPHFIGXRSVITXLME

Der p 5
MKFIIAFFVATLAVMTVSGEDKKHDYQNEFDFLLMERIHEQIKKGELALFYLQ
EQINHFEEKPTKEMKDKIVAEMDTIIAMIDGVRGVLDRLMQRKDLDIFEQYN
LEMAKKSGDILERDLKKEEARVKKIEV

Der p 6
AIGXQPAAEAEAPFQISLMK

Der p 7
MMKLLLIAAAAFVAVSADPIHYDKITEEINKAVDEAVAAIEKSETFDPMKVP
DHSDKFERHIGIIDLKGELDMRNIQVRGLKQMKRVGDANVKSEDGVVKAHL
LVGVHDDVVSMEYDLAYKLGDLHPNTHVISDIQDFVVELSLEVSEEGNMTLT
SFEVRQFANVVNHIGGLSILDPIFAVLSDVLTAIFQDTVRAEMTKVLAPAFKK
ELERNNQ

Der p9
IVGGSNASPGDAVYQIAL

Dermatophagoides farinae

Der f 1
MKFVLAIASLLVLTVYARPASIKTFEFKKAFNKNYATVEEEEVARKNFLESLK

YVEANKGAINHLSDLSLDEFKNRYLMSAEAFEQLKTQFDLNAETSACRINSV
NVPSELDLRSLRTVTPIRMQGGCGSCWAFSGVAATESAYLAYRNTSLDLSEQ
ELVDCASQHGCHGDTIPRGIEYIQQNGVVEERSYPYVAREQRCRRPNSQHYG
ISNYCQIYPPDVKQIREALTQTHTAIAVIIGIKDLRAFQHYDGRTIIQHDNGYQP
NYHAVNIVGYGSTQGDDYWIVRNSWDTTWGDSGYGYFQAGNNLMMIEQY
PYVVIM

Der f 2

MISKILCLSLLVAAVVADQVDVKDCANNEIKKVMVDGCHGSDPCIIHRGKPF
TLEALFDANQNTKTAKIEIKASLDGLEIDVPGIDTNACHFMKCPLVKGQQYDI
KYTWNVPKIAPKSENVVVTVKLIGDNGVLACAIATHGKIRD

Der f 3

MMILTIVVLLAANILATPILPSSPNATIVGGVKAQAGDCPYQISLQSSSHFCGG
SILDEYWILTAAHCVNGQSAKKLSIRYNTLKHASGGEKIQVAEIYQHENYDS
MTIDNDVALIKLKTPMTLDQTNAKPVPLPAQGSDVKVGDKIRVSGWGYLQE
GSYSLPSELQRVDIDVVSREQCDQLYSKAGADVSENMICGGDVANGGVDSC
QGDSGGPVVDVATKQIVGIVSWGYGCARKGYPGVYTRVGNFVDWIESKRS
Q

Der f 4

AVGGQDADLAEAPFQISLLK

Der f 7

MMKFLLIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAIEQSETIDPMKVPDH
ADKFERHVGIVDFKGELAMRNIEARGLKQMKRQGDANVKGEEGIVKAHLLI
GVHDDIVSMEYDLAYKLGDLHPTTHVISDIQDFVVALSLEISDEGNITMTSFE
VRQFANVVNHIGGLSILDPIFGVLSDVLTAIFQDTVRKEMTKVLAPAFKRELE
KN

Additional mite allergen sequences (NCBI entrez accession):

1170095; 1359436; 2440053; 666007; 487661; 1545803; 84702; 84699; 625532; 404370; 1091577; 1460058; 7413; 9072; 387592.

Cat

Felis sequences (NCBI entrez accession):

539716; 539715; 423193; 423192; 423191; 423190; 1364213; 1364212; 395407; 163827; 163823; 163825; 1169665; 232086; 1169666.

Latex

Hevea sequences:

Hev b 1

MAEDEDNQQGQGEGLKYLGFVQDAATYAVTTFSNVYLFAKDKSGPLQPGV
DIIEGPVKNVAVPLYNRFSYIPNGALKFVDSTVVASVTIIDRSLPPIVKDASIQV
VSAIRAAPEAARSLASSLPGQTKILAKVFYGEN

Hev b 3

MAEEVEEERLKYLDFVRAAGVYAVDSFSTLYLYAKDISGPLKPGVDTIENVV
KTVVTPVYYIPLEAVKFVDKTVDVSVTSLDGVVPPVIKQVSAQTYSVAQDAP
RIVLDVASSVFNTGVQEGAKALYANLEPKAEQYAVITWRALNKLPLVPQVA
NVVVPTAVYFSEKYNDVVRGTTEQGYRVSSYLPLLPTEKITKVFGDEAS

Additional Hevea sequences (NCBI entrez accession):

3319923; 3319921; 3087805; 1493836; 1480457; 1223884; 3452147; 3451147; 1916805; 232267; 123335; 2501578; 3319662; 3288200; 1942537; 2392631; 2392630; 1421554; 1311006; 494093; 3183706; 3172534; 283243; 1170248; 1708278; 1706547; 464775; 2661042; 231586; 123337; 116359; 123062; 2213877; 542013; 2144920; 1070656; 2129914; 2129913; 2129912; 100135; 82026; 1076559;

82028; 82027; 282933; 280399; 100138; 1086972; 108697; 1086976; 1086978; 1086978; 1086976; 1086974; 1086972; 913758; 913757; 913756; 234388; 1092500; 228691; 1177405; 18839; 18837; 18835; 18833; 18831; 1209317; 1184668; 168217; 168215; 168213; 168211; 168209; 348137.

Rye grass

Lolium sequences:

126385 Lol p 1

MASSSSVLLVVALFAVFLGSAHGIAKVPPGPNITAEYGDKWLDAKSTWYGK
PTGAGPKDNGGACGYKNVDKAPFNGMTGCGNTPIFKDGRGCGSCFEIKCTK
PESCSGEAVTVTITDDNEEPIAPYHFDLSGHAFGSMAKKGEEQNVRSAGELEL
QFRRVKCKYPDDTKPTFHVEKASNPNYLAILVKYVDGDGDVVAVDIKEKGK
DKWIELKESWGAVWRIDTPDKLTGPFTVRYTTEGGTKSEFEDVIPEGWKADT
SYSAK

126386 Lol p 2a

AAPVEFTVEKGSDEKNLALSIKYNKEGDSMAEVELKEHGSNEWLALKKNGD
GVWEIKSDKPLKGPFNFRFVSEKGMRNVFDDVVPADFKVGTTYKPE

126387 Lol p 3

TKVDLTVEKGSDAKTLVLNIKYTRPGDTLAEVELRQHGSEEWEPMTKKGNL
WEVKSAKPLTGPMNFRFLSKGGMKNVFDEVIPTAFTVGKTYTPEYN

2498581 Lol p 5a

MAVQKYTVALFLRRGPRGGPGRSYAADAGYTPAAAATPATPAATPAGGWR
EGDDRRAEAAGGRQRLASRQPWPPLPTPLRRTSSRSSRPPSPSPPRASSPTSA
AKAPGLIPKLDTAYDVAYKAAEAHPRGQVRRLRHCPHRSLRVIAGALEVHA
VKPATEEVLAAKIPTGELQIVDKIDAAFKIAATAANAAPTNDKFTVFESAFNK
ALNECTGGAMRPTSSSPPSRPRSSRPTPPPSPAAPEVKYAVFEAALTKAITAM
TQAQKAGKPAAAAATAAATVATAAATAAAVLPPPLLVVQSLISLLIYY

2498582 Lol p 5b

MAVQKHTVALFLAVALVAGPAASYAADAGYAPATPATPAAPATAATPATP
ATPATPAAVPSGKATTEEQKLIEKINAGFKAAVAAAAVVPPADKYKTFVETF
GTATNKAFVEGLASGYADQSKNQLTSKLDAALKLAYEAAQGATPEAKYDA
YVATLTEALRVIAGTLEVHAVKPAAEEVKVGAIPAAEVQLIDKVDAAYRTA
ATAANAAPANDKFTVFENTFNNAIKVSLGAAYDSYKFIPTLVAAVKQAYAA
KQATAPEVKYTVSETALKKAVTAMSEAEKEATPAAAATATPTPAAATATAT
PAAAYATATPAAATATATPAAATATPAAAGGYKV

455288 Lol p isoform 9

MAVQKHTVALFLAVALVAGPAASYAADAGYAPATPATPAAPATAATPATP
ATPATPAAVPSGKATTEEQKLIEKINAGFKAAVAAAAVVPPADKYKTFVETF
GTATNKAFVEGLASGYADQSKNQLTSKLDAALKLAYEAAQGATPEAKYDA
YVATLTEALRVIAGTLEVHAVKPAAEEVKVGAIPAAEVQLIDKVDAAYRTA
ATAANAAPANDKFTVFENTFNNAIKVSLGAAYDSYKFIPTLVAAVKQAYAA
KQATAPEVKYTVSETALKKAVTAMSEAEKEATPAAAATATPTPAAATATAT
PAAAYATATPAAATATATPAAATATPAAAGGYKV

1582249 Lol p 11

DKGPGFVVTGRVYCDPCRAGFETNVSHNVEGATVAVDCRPFDGGESKLKAE
ATTDKDGWYKIEIDQDHQEEICEVVLAKSPDKSCSEIEEFRDRARVPLTSNXG
IKQQGIRYANPIAFFRKEPLKECGGILQAY

Additional Lolium sequences (NCBI entrez accession):

135480; 417103; 687261; 687259; 1771355; 2388662; 631955; 542131; 542130; 542129; 100636; 626029; 542132; 320616; 320615; 320614; 100638; 100634; 82450; 626028; 100639; 283345; 542133; 1771353; 1763163; 1040877; 1040875; 250525; 551047; 515377; 510911; 939932; 439950; 2718; 168316; 168314; 485371; 2388664; 2832717; 2828273; 548867.

Olive tree

Olive sequences

416610 Ole e 1

EDIPQPPVSQFHIQGQVYCDTCRAGFITELSEFIPGASLRLQCKDKENGDVTFT
EVGYTRAEGLYSMLVERDHKNEFCEITLISSGRKDCNEIPTEGWAKPSLKFKL
NTVNGTTRTVNPLGFFKKEALPKCAQVYNKLGMYPPNM

Parietaria

Parietaria sequences:

2497750 Par j P2

MRTVSMAALVVIAAALAWTSSAEPAPAPAPGEEACGKVVQDIMPCLHFVKG
EEKEPSKECCSGTKKLSEEVKTTEQKREACKCIVRATKGISGIKNELVAEVPK
KCDIKTTLPPITADFDCSKIQSTIFRGYY

1352506 Par j P5

MVRALMPCLPFVQGKEKEPSKGCCSGAKRLDGETKTGPQRVHACECIQTAM
KTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPVSLRH
GPVTGPSDPAHKARLERPQIRVPPPAPEKA

1532056 Par j P8

MRTVSMAALVVIAAALAWTSSAELASAPAPGEGPCGKVVHHIMPCLKFVKG
EEKEPSKSCCSGTKKLSEEVKTTEQKREACKCIVAATKGISGIKNELVAEVPK
KCGITTTLPPITADFDCSKIESTIFRGYY

1532058 Par j P9

MRTVSAPSAVALVVIVAAGLAWTSLASVAPPAPAPGSEETCGTVVRALMPC
LPFVQGKEKEPSKGCCSGAKRLDGETKTGLQRVHACECIQTAMKTYSDIDGK
LVSEVPKHCGIVDSKLPPIDVNMDCKTLGVVPRQPQLPVSLRHGPVTGPSDPA

HKARLERPQIRVPPPAPEKA

2497749 Par j P9

MRTVSARSSVALVVIVAAVLVWTSSASVAPAPAPGSEETCGTVVGALMPCL
PFVQGKEKEPSKGCCSGAKRLDGETKTGPQRVHACECIQTAMKTYSDIDGKL
VSEVPKHCGIVDSKLPPIDVNMDCKTLGVLHYKGN

1086003 Par j 1

MVRALMPCLPFVQGKEKEPSKGCCSGAKRLDGETKTGPQRVHACECIQTAM
KTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPVSLRH
GPVTGPSRSRPPTKHGWRDPRLEFRPPHRKKPNPAFSTLG

Additional Parietaria sequences (NCBI entrez accession):

543659; 1836011; 1836010; 1311513; 1311512; 1311511; 1311510; 1311509; 240971.

Timothy grass

Phleum sequences:

Phl p 1

MASSSSVLLVVVLFAVFLGSAYGIPKVPPGPNITATYGDKWLDAKSTWYGKP
TGAGPKDNGGACGYKDVDKPPFSGMTGCGNTPIFKSGRGCGSCFEIKCTKPE
ACSGEPVVVHITDDNEEPIAPYHFDLSGHAFGAMAKKGDEQKLRSAGELELQ
FRRVKCKYPEGTKVTFHVEKGSNPNYLALLVKYVNGDGDVVAVDIKEKGK
DKWIELKESWGAIWRIDTPDKLTGPFTVRYTTEGGTKTEAEDVIPEGWKADT
SYESK

Phl p 1

MASSSSVLLVVALFAVFLGSAHGIPKVPPGPNITATYGDKWLDAKSTWYGKP
TAAGPKDNGGACGYKDVDKPPFSGMTGCGNTPIFKSGRGCGSCFEIKCTKPE

ACSGEPVVVHITDDNEEPIAAYHFDLSGIAFGSMAKKGDEQKLRSAGEVEIQF
RRVKCKYPEGTKVTFHVEKGSNPNYLALLVKFSGDGDVVAVDIKEKGKDK
WIALKESWGAIWRIDTPEVLKGPFTVRYTTEGGTKARAKDVIPEGWKADTA
YESK

Phlp 2

MSMASSSSSSLLAMAVLAALFAGAWCVPKVTFTVEKGSNEKHLAVLVKYE
GDTMAEVELREHGSDEWVAMTKGEGGVWTFDSEEPLQGPFNFRFLTEKGM
KNVFDDVVPEKYTIGATYAPEE

Phl p 5

ADLGYGGPATPAAPAEAAPAGKATTEEQKLIEKINDGFKAALAAAAGVPPA
DKYKTFVATFGAASNKAFAEGLSAEPKGAAESSSKAALTSKLDAAYKLAYK
TAEGATPEAKYDAYVATLSEALRIIAGTLEVHAVKPAAEEVKVIPAGELQVIE
KVDSAFKVAATAANAAPANDKFTVFEAAFNNAIKASTGGAYESYKFIPALEA
AVKQAYAATVATAPEVKYTVFETALKKAFTAMSEAQKAAKPATEATATAT
AAVGAATGAATAATGGYKV

Phl p 5

ADLGYGGPATPAAPAEAAPAGKATTEEQKLIEKINDGFKAALAAAAGVPPA
DKYKTFVATFGAASNKAFAEGLSAEPKGAAESSSKAALTSKLDAAYKLAYK
TAEGATPEAKYDAYVATLSEALRIIAGTLEVHAVKPAAEEVKVIPAGELQVIE
KVDSAFKVAATAANAAPANDKFTVFEAAFNNAIKASTGGAYESYKFIPALEA
AVKQAYAATVATAPEVKYTVFETALKKAITAMSEAQKAAKPATEATATAT
AAVGAATGAATAATGGYKV

Phl p 5b

AAAAVPRRGPRGGPGRSYTADAGYAPATPAAAGAAAGKATTEEQKLIEDIN
VGFKAAVAAAASVPAADKFKTFEAAFTSSSKAAAAKAPGLVPKLDAAYSVA
YKAAVGATPEAKFDSFVASLTEALRVIAGALEVHAVKPVTEEPGMAKIPAGE
LQIIDKIDAAFKVAATAAATAPADDKFTVFEAAFNKAIKESTGGAYDTYKCIP
SLEAAVKQAYAATVAAAPQVKYAVFEAALTKAITAMSEVQKVSQPATGAA
TVAAGAATTAAGAASGAATVAAGGYKV

Phl p 5a

ADLGYGPATPAAPAAGYTPATPAAPAGADAAGKATTEEQKLIEKINAGFKA
ALAGAGVQPADKYRTFVATFGPASNKAFAEGLSGEPKGAAESSSKAALTSK
LDAAYKLAYKTAEGATPEAKYDAYVATLSEALRIIAGTLEVHAVKPAAEEV
KVIPAGELQVIEKVDAAFKVAATAANAAPANDKFTVFEAAFNDEIKASTGGA
YESYKFIPALEAAVKQAYAATVATAPEVKYTVFETALKKAITAMSEAQKAA
KPAAAATATATAAVGAATGAATAATGGYKV

Phl p 5

MAVQKYTVALFLAVALVAGPAASYAADAGYAPATPAAAGAEAGKATTEEQ
KLIEDINVGFKAAVAAAASVPAADKFKTFEAAFTSSSKAATAKAPGLVPKLD
AAYSVSYKAAVGATPEAKFDSFVASLTEALRVIAGALEVHAVKPVTEEPGM
AKIPAGELQIIDKIDAAFKVAATAAATAPADTVFEAAFNKAIKESTGGAYDTY
KCIPSLEAAVKQAYAATVAAAPQVKYAVFEAALTKAITAMSEVQKVSQPAT
GAATVAAGAATTAAGAASGAATVAAGGYKV

Phl p 5

MAVQKYTVALFLAVALVAGPAASYAADAGYAPATPAAAGAEAGKATTEEQ
KLIEDINVGFKAAVAAAASVPAADKFKTFEAAFTSSSKAATAKAPGLVPKLD
AAYSVAYKAAVGATPEAKFDSFVASLTEALRVIAGALEVHAVKPVTEDPAW
PKIPAGELQIIDKIDAAFKVAATAAATAPADDKFTVFEAAFNKAIKESTGGAY
DTYKCIPSLEAAVKQAYAATVAAAPQVKYAVFEAALTKAITAMSEVQKVSQ
PATGAATVAAGAATTATGAASGAATVAAGGYKV

Phl p 5

ADAGYAPATPAAAGAEAGKATTEEQKLIEDINVGFKAAVAAAASVPAADKF
KTFEAAFTSSSKAATAKAPGLVPKLDAAYSVAYKAAVGATPEAKFDSFVAS
LTEALRVIAGALEVHAVKPVTEEPGMAKIPAGELQIIDKIDAAFKVAATAAAT
APADDKFTVFEAAFNKAIKESTGGAYDTYKCIPSLEAAVKQAYAATVAAAP
QVKYAVFEAALTKAITAMSEVQKVSQPATGAATVAAGAATTAAGAASGAA
TVAAGGYKV

Phl p 5

SVKRSNGSAEVHRGAVPRRGPRGGPGRSYAADAGYAPATPAAAGAEAGKA
TTEEQKLIEDINVGFKAAVAAAASVPAADKFKTFEAAFTSSSKAATAKAPGL
VPKLDAAYSVAYKAAVGATPEAKFDSFVASLTEALRVIAGALEVHAVKPVT
EEPGMAKIPAGELQIIDKIDAAFKVAATAAATAPADDKFTVFEAAFNKAIKES
TGGAYDTYKCIPSLEAAVKQAYAATVAAAPQVKYAVFEAALTKAITAMSEV
QKVSQPATGAATVAAGAATTAAGAASGAATVAAGGYKV

Phl p 5

MAVHQYTVALFLAVALVAGPAGSYAADLGYGPATPAAPAAGYTPATPAAP
AGAEPAGKATTEEQKLIEKINAGFKAALAAAAGVPPADKYRTFVATFGAAS
NKAFAEGLSGEPKGAAESSSKAALTSKLDAAYKLAYKTAEGATPEAKYDAY
VATVSEALRIIAGTLEVHAVKPAAEEVKVIPAGELQVIEKVDAAFKVAATAA
NAAPANDKFTVFEAAFNDAIKASTGGAYESYKFIPALEAAVKQAYAATVAT
APEVKYTVFETALKKAITAMSEAQKAAKPAAAATATATAAVGAATGAATA
ATGGYKV

Phl p 5

ADLGYGGPATPAAPAEAAPAGKATTEEQKLIEKINDGFKAALAAAAGVPPA
DKYKTFVATFGAASNKAFAEGLSAEPKGAAESSSKAALTSKLDAAYKLAYK
TAEGATPEAKYDAYVATLSEALRIIAGTLEVHAVKPAAEEVKVIPAGELQVIE
KVDSAFKVAATAANAAPANDKFTVFEAAFNNAIKASTGGAYESYKFIPALEA

AVKQAYAATVATAPEVKYTVFETALKKAFTAMSEAQKAAKPATEATATAT
AAVGAATGAATAATGGYKV

Phl p5b

AAAAVPRRGPRGGPGRSYTADAGYAPATPAAAGAAAGKATTEEQKLIEDIN
VGFKAAVAAAASVPAADKFKTFEAAFTSSSKAAAAKAPGLVPKLDAAYSVA
YKAAVGATPEAKFDSFVASLTEALRVIAGALEVHAVKPVTEEPGMAKIPAGE
LQIIDKIDAAFKVAATAAATAPADDKFTVFEAAFNKAIKESTGGAYDTYKCIP
SLEAAVKQAYAATVAAAPQVKYAVFEAALTKAITAMSEVQKVSQPATGAA
TVAAGAATTAAGAASGAATVAAGGYKV

Phl p5a

ADLGYGPATPAAPAAGYTPATPAAPAGADAAGKATTEEQKLIEKINAGFKA
ALAGAGVQPADKYRTFVATFGPASNKAFAEGLSGEPKGAAESSSKAALTSK
LDAAYKLAYKTAEGATPEAKYDAYVATLSEALRIIAGTLEVHAVKPAAEEV
KVIPAGELQVIEKVDAAFKVAATAANAAPANDKFTVFEAAFNDEIKASTGGA
YESYKFIPALEAAVKQAYAATVATAPEVKYTVFETALKKAITAMSEAQKAA
KPAAAATATATAAVGAATGAATAATGGYKV

Phl p 5

AVPRRGPRGGPGRSYAADAGYAPATPAAAGAEAGKATTEEQKLIEDINVGF
KAAVAAAASVPAGDKFKTFEAAFTSSSKAATAKAPGLVPKLDAAYSVAYKA
AVGATPEAKFDSFVASLTEALRVIAGALEVHAVKPVTEEPGMAKIPAGELQII
DKIDAAFKVAATAAATAPADDKFTVFEAAFNKAIKESTGGAYDTYKCIPSLE
AAVKQAYAATVAAAPQVKYAVFEAALTKAITAMSEVQKVSQPATGAATVA
AGAATTATGAASGAATVAAGGYKV

Phl p 5b

MAVPRRGPRGGPGRSYTADAGYAPATPAAAGAAAGKATTEEQKLIEDINVG
FKAAVAARQRPAADKFKTFEAASPRHPRPLRQGAGLVPKLDAAYSVAYKAA
VGATPEAKFDSFVASLTEALRVIAGALEVHAVKPVTEEPGMAKIPAGELQIID

KIDAAFKVAATAAATAPADDKFTVFEAAFNKAIKESTGGAYDTYKCIPSLEA
AVKQAYAATVAAAAEVKYAVFEAALTKAITAMSEVQKVSQPATGAATVAA
GAATTAAGAASGAATVAAGGYKV

Phl p 5
MAVHQYTVALFLAVALVAGPAASYAADLGYGPATPAAPAAGYTPATPAAP
AEAAPAGKATTEEQKLIEKINAGFKAALAAAAGVQPADKYRTFVATFGAAS
NKAFAEGLSGEPKGAAESSSKAALTSKLDAAYKLAYKTAEGATPEAKYDAY
VATLSEALRIIAGTLEVHAVKPAAEEVKVIPAGELQVIEKVDAAFKVAATAA
NAAPANDKFTVFEAAFNDAIKASTGGAYESYKFIPALEAAVKQAYAATVAT
APEVKYTVFETALKKAITAMSEAQKAAKPAAAATATATAAVGAATGAATA
ATGGYKV

Phl p 5
EAPAGKATTEEQKLIEKINAGFKAALARRLQPADKYRTFVATFGPASNKAFA
EGLSGEPKGAAESSSKAALTSKLDAAYKLAYKTAEGATPEAKYDAYVATLS
EALRIIAGTLEVHAVKPAAEEVKVIPAAELQVIEKVDAAFKVAATAANAAPA
NDKFTVFEAAFNDEIKASTGGAYESYKFIPALEAAVKQAYAATVATAPEVKY
TVFETALKKAITAMSEAQKAAKPPPLPPPPQPPPLAATGAATAATGGYKV

Phl p 5
MAVHQYTVALFLAVALVAGPAASYAADLGYGPATPAAPAAGYTPATPAAP
AEAAPAGKATTEEQKLIEKINAGFKAALAAAAGVQPADKYRTFVATFGAAS
NKAFAEGLSGEPKGAAESSSKAALTSKLDAAYKLAYKTAEGATPEAKYDAY
VATLSEALRIIAGTLEVHAVKPAAEEVKVIPAGELQVIEKVDAAFKVAATAA
NAAPANDKFTVFEAAFNDAIKASTGGAYESYKFIPALEAAVKQAYAATVAT
APEVKYTVFETALKKAITAMSEAQKAAKPAAAATATATAAVGAATGAATA
ATGGYKV

Phl p 5b
MAVPRRGPRGGPGRSYTADAGYAPATPAAAGAAAGKATTEEQKLIEDINVG

FKAAVAARQRPAADKFKTFEAASPRHPRPLRQGAGLVPKLDAAYSVAYKAA
VGATPEAKFDSFVASLTEALRVIAGALEVHAVKPVTEEPGMAKIPAGELQIID
KIDAAFKVAATAAATAPADDKFTVFEAAFNKAIKESTGGAYDTYKCIPSLEA
AVKQAYAATVAAAAEVKYAVFEAALTKAITAMSEVQKVSQPATGAATVAA
GAATTAAGAASGAATVAAGGYKV

Phl p 5a

ADLGYGPATPAAPAAGYTPATPAAPAGADAAGKATTEEQKLIEKINAGFKA
ALAGAGVQPADKYRTFVATFGPASNKAFAEGLSGEPKGAAESSSKAALTSK
LDAAYKLAYKTAEGATPEAKYDAYVATLSEALRIIAGTLEVHAVKPAAEEV
KVIPAGELQVIEKVDAAFKVAATAANAAPANDKFTVFEAAFNDEIKASTGGA
YESYKFIPALEAAVKQAYAATVATAPEVKYTVFETALKKAITAMSEAQKAA
KPPPLPPPPQPPPLAATGAATAATGGYKV

Phl p 5

MAVHQYTVALFLAVALVAGPAASYAADLGYGPATPAAPAAGYTPATPAAP
AEAAPAGKATTEEQKLIEKINAGFKAALAAAAGVQPADKYRTFVATFGAAS
NKAFAEGLSGEPKGAAESSSKAALTSKLDAAYKLAYKTAEGATPEAKYDAY
VATLSEALRIIAGTLEVHAVKPAAEEVKVIPAGELQVIEKVDAAFKVAATAA
NAAPANDKFTVFEAAFNDAIKASTGGAYESYKFIPALEAAVKQAYAATVAT
APEVKYTVFETALKKAITAMSEAQKAAKPAAAATATATAAVGAATGAATA
ATGGYKV

Phl p 6

MAAHKFMVAMFLAVAVVLGLATSPTAEGGKATTEEQKLIEDVNASFRAAM
ATTANVPPADKYKTFEAAFTVSSKRNLADAVSKAPQLVPKLDEVYNAAYNA
ADHAAPEDKYEAFVLHFSEALRIIAGTPEVHAVKPGA

Phl p 6

SKAPQLVPKLDEVYNAAYNAADHAAPEDKYEAFVLHFSEALHIIAGTPEVHA
VKPGA

Phl p 6

ADKYKTFEAAFTVSSKRNLADAVSKAPQLVPKLDEVYNAAYNAADHAAPE
DKYEAFVLHFSEALHIIAGTPEVHAVKPGA

Phl p 6

TEEQKLIEDVNASFRAAMATTANVPPADKYKTLEAAFTVSSKRNLADAVSK
APQLVPKLDEVYNAAYNAADHAAPEDKYEAFVLHFSEALRIIAGTPEVHAVK
PGA

Phl p 6

MAAHKFMVAMFLAVAVVLGLATSPTAEGGKATTEEQKLIEDINASFRAAMA
TTANVPPADKYKTFEAAFTVSSKRNLADAVSKAPQLVPKLDEVYNAAYNAA
DHAAPEDKYEAFVLHFSEALHIIAGTPEVHAVKPGA

Phl p 6

MVAMFLAVAVVLGLATSPTAEGGKATTEEQKLIEDVNASFRAAMATTANVP
PADKYKTFEAAFTVSSKRNLADAVSKAPQLVPKLDEVYNAAYNAADHAAP
EDKYEAFVLHFSEALRIIAGTPEVHAVKPGA

Phl p 7

MADDMERIFKRFDTNGDGKISLSELTDALRTLGSTSADEVQRMMAEIDTDGD
GFIDFNEFISFCNANPGLMKDVAKVF

Phl p 11

MSWQTYVDEHLMCEIEGHHLASAAILGHDGTVWAQSADFPQFKPEEITGIM
KDFDEPGHLAPTGMFVAGAKYMVIQGEPGRVIRGKKGAGGITIKKTGQALV
VGIYDEPMTPGQCNMVVERLGDYLVEQGM

Additional Phleum sequences (NCBI entrez accession):

458878; 548863; 2529314; 2529308; 2415702; 2415700; 2415698; 542168; 542167; 626037; 542169; 541814; 542171; 253337; 253336; 453976; 439960 .

Wasp (and related)

Vespula sequences:

465054 ALLERGEN VES V 5

MEISGLVYLIIIVTIIDLPYGKANNYCKIKCLKGGVHTACKYGSLKPNCGNKV
VVSYGLTKQEKQDILKEHNDFRQKIARGLETRGNPGPQPPAKNMKNLVWND
ELAYVAQVWANQCQYGHDTCRDVAKYQVGQNVALTGSTAAKYDDPVKLV
KMWEDEVKDYNPKKKFSGNDFLKTGHYTQMVWANTKEVGCGSIKYIQEK
WHKHYLVCNYGPSGNFMNEELYQTK

1709545 ALLERGEN VES M 1

GPKCPFNSDTVSIIIETRENRNRDLYTLQTLQNHPEFKKKTITRPVVFITHGFTS
SASEKNFINLAKALVDKDNYMVISIDWQTAACTNEYPGLKYAYYPTAASNT
RLVGQYIATITQKLVKDYKISMANIRLIGHSLGAHVSGFAGKRVQELKLGKYS
EIIGLDPARPSFDSNHCSERLCETDAEYVQIIHTSNYLGTEKILGTVDFYMNNG
KNNPGCGRFFSEVCSHTRAVIYMAECIKHECCLIGIPRSKSSQPISRCTKQECV
CVGLNAKKYPSRGSFYVPVESTAPFCNNKGKII

1352699 ALLERGEN VES V 1

MEENMNLKYLLLFVYFVQVLNCCYGHGDPLSYELDRGPKCPFNSDTVSIIIET
RENRNRDLYTLQTLQNHPEFKKKTITRPVVFITHGFTSSASETNFINLAKALVD
KDNYMVISIDWQTAACTNEAAGLKYLYYPTAARNTRLVGQYIATITQKLVK
HYKISMANIRLIGHSLGAHASGFAGKKVQELKLGKYSEIIGLDPARPSFDSNH
CSERLCETDAEYVQIIHTSNYLGTEKTLGTVDFYMNNGKNQPGCGRFFSEVC
SHSRAVIYMAECIKHECCLIGIPKSKSSQPISSCTKQECVCVGLNAKKYPSRGS
FYVPVESTAPFCNNKGKII

1346323 ALLERGEN VES V 2

SERPKRVFNIYWNVPTFMCHQYDLYFDEVTNFNIKRNSKDDFQGDKIAIFYD
PGEFPALLSLKDGKYKKRNGGVPQEGNITIHLQKFIENLDKIYPNRNFSGIGVI
DFERWRPIFRQNWGNMKIHKNFSIDLVRNEHPTWNKKMIELEASKRFEKYA
RFFMEETLKLAKKTRKQADWGYYGYPYCFNMSPNNLVPECDVTAMHENDK
MSWLFNNQNVLLPSVYVRQELTPDQRIGLVQGRVKEAVRISNNLKHSPKVLS
YWWYVYQDETNTFLTETDVKKTFQEIVINGGDGIIIWGSSSDVNSLSKCKRL
QDYLLTVLGPIAINVTEAVN

549194 ALLERGEN VES VI
5KVNYCKIKCLKGGVHTACKYGTSTKPNCGKMVVKAYGLTEAEKQEILKVH
NDFRQKVAKGLETRGNPGPQPPAKNMNNLVWNDELANIAQVWASQCNYG
HDTCKDTEKYPVGQNIAKRSTTAALFDSPGKLVKMWENEVKDFNPNIEWSK
NNLKKTGHYTQMVWAKTKEIGCGSVKYVKDEWYTHYLVCNYGPSGNFRN
EKLYEKK

Additional vespula sequences (NCBI entrez accession):
549193; 549192; 549191; 549190; 5491104; 117414; 126761; 69576; 625255; 6271104; 627188; 627187; 482382; 112561; 627186; 627185; 1923233; 1047645; 1047647; 745570; 225764; 162551.

<u>Tree allergen sequences (mainly birch) sequences:</u>

114922 Bet v 1
MGVFNYETETTSVIPAARLFKAFILDGDNLFPKVAPQAISSVENIEGNGGPGTI
KKISFPEGFPFKYVKDRVDEVDHTNFKYNYSVIEGGPIGDTLEKISNEIKIVAT
PDGGSILKISNKYHTKGDHEVKAEQVKASKEMGETLLRAVESYLLAHSDAY
N

130975 Bet v 2
MSWQTYVDEHLMCDIDGQASNSLASAIVGHDGSVWAQSSSFPQFKPQEITGI
MKDFEEPGHLAPTGLHLGGIKYMVIQGEAGAVIRGKKGSGGITIKKTGQALV

FGIYEEPVTPGQCNMVVERLGDYLIDQGL

1168696 Bet v 3

MPCSTEAMEKAGHGHASTPRKRSLSNSSFRLRSESLNTLRLRRIFDLFDKNSD
GIITVDELSRALNLLGLETDLSELESTVKSFTREGNIGLQFEDFISLHQSLNDSY
FAYGGEDEDDNEEDMRKSILSQEEADSFGGFKVFDEDGDGYISARELQMVL
GKLGFSEGSEIDRVEKMIVSVDSNRDGRVDFFEFKDMMRSVLVRSS

809536 Bet v 4

MADDHPQDKAERERIFKRFDANGDGKISAAELGEALKTLGSITPDEVKHMM
AEIDTDGDGFISFQEFTDFGRANRGLLKDVAKIF

543675 Que a I - Quercus alba=oak trees (fragment)

GVFTXESQETSVIAPAXLFKALFL

543509 Car b I - Carpinus betulus=hornbeam trees (fragment)

GVFNYEAETPSVIPAARLFKSYVLDGDKLIPKVAPQAIXK

543491 Aln g I - Alnus glutinosa=alder trees (fragment)

GVFNYEAETPSVIPAARLFKAFILDGDKLLPKVAPEAVSSVENI

1204056 Rubisco

VQCMQVWPPLGLKKFETLSYLPPLSSEQLAKEVDYLLRKNLIPCLEFELEHGF
VYREHNRSPGYYDGRYWTMWKLPMFGCNDSSQVLKELEECKKAYPSAFIRI
IGFDDK

Additional tree allergen sequences (NCBI entrez accession number):

131919; 128193; 585564; 1942360; 2554672; 2392209; 2414158; 1321728;
1321726; 1321724; 1321722; 1321720; 1321718; 1321716; 1321714; 1321712;
3015520; 2935416; 464576; 1705843; 1168701; 1168710; 1168709; 1168708;

1168707; 1168706; 1168705; 1168704; 1168703; 1168702; 1842188; 2564228; 2564226; 2564224; 2564222; 2564220; 2051993; 18131041; 15368104; 534910; 534900; 5341048; 1340000; 1339998; 2149808; 66207; 2129477; 1076249; 1076247; 629480; 481805; 81443; 1361968; 1361967; 1361966; 1361965; 1361964; 1361963; 1361962; 1361961; 1361960; 1361959; 320546; 629483 ; 629482; 629481; 541804; 320545; 81444; 541814:; 629484; 474911; 452742; 1834387; 298737; 298736; 1584322; 1584321; 584320; 1542873; 1542871; 1542869; 1542867; 1542865; 1542863; 1542861; 1542859; 1542857; 1483232; 1483230; 1483228; 558561; 551640; 488605; 452746; 452744; 452740; 452738; 452736; 452734; 452732; 452730; 452728; 450885; 17938; 17927; 17925; 17921; 297538; 510951; 2104331; 2104329; 166953 .

Peanut

Peanut sequences

1168391 Ara h 1

MRGRVSPLMLLLGILVLASVSATHAKSSPYQKKTENPCAQRCLQSCQQEPDD
LKQKACESRCTKLEYDPRCVYDPRGHTGTTNQRSPPGERTRGRQPGDYDDD
RRQPRREEGGRWGPAGPREREREEDWRQPREDWRRPSHQQPRKIRPEGREG
EQEWGTPGSHVREETSRNNPFYFPSRRFSTRYGNQNGRIRVLQRFDQRSRQF
QNLQNHRIVQIEAKPNTLVLPKHADADNILVIQQGQATVTVANGNNRKSFNL
DEGHALRIPSGFISYILNRHDNQNLRVAKISMPVNTPGQFEDFFPASSRDQSSY
LQGFSRNTLEAAFNAEFNEIRRVLLEENAGGEQEERGQRRWSTRSSENNEGVI
VKVSKEHVEELTKHAKSVSKKGSEEEGDITNPINLREGEPDLSNNFGKLFEVK
PDKKNPQLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKGTGNLELV
AVRKEQQQRGRREEEEDEDEEEEGSNREVRRYTARLKEGDVFIMPAAHPVAI
NASSELHLLGFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKL
IKNQKESHFVSARPQSQSQSPSSPEKESPEKEDQEEENQGGKGPLLSILKAFN

Ragweed

Ambrosia sequences

113478 Amb a 1

MGIKHCCYILYFTLALVTLLQPVRSAEDLQQILPSANETRSLTTCGTYNIIDGC
WRGKADWAENRKALADCAQGFAKGTIGGKDGDIYTVTSELDDDVANPKEG
TLRFGAAQNRPLWIIFARDMVIRLDRELAINNDKTIDGRGAKVEIINAGFAIYN
VKNIIIHNIIMHDIVVNPGGLIKSHDGPPVPRKGSDGDAIGISGGSQIWIDHCSLS
KAVDGLIDAKHGSTHFTVSNCLFTQHQYLLLFWDFDERGMLCTVAFNKFTD
NVDQRMPNLRHGFVQVVNNNYERWGSYALGGSAGPTILSQGNRFLASDIKK
EVVGRYGESAMSESINWNWRSYMDVFENGAIFVPSGVDPVLTPEQNAGMIP
AEPGEAVLRLTSSAGVLSCQPGAPC

113479 Amb a 2

MGIKHCCYILYFTLALVTLVQAGRLGEEVDILPSPNDTRRSLQGCEAHNIIDK
CWRCKPDWAENRQALGNCAQGFGKATHGGKWGDIYMVTSDQDDDVVNP
KEGTLRFGATQDRPLWIIFQRDMIIYLQQEMVVTSDKTIDGRGAKVELVYGGI
TLMNVKNVIIHNIDIHDVRVLPGGRIKSNGGPAIPRHQSDGDAIHVTGSSDIWI
DHCTLSKSFDGLVDVNWGSTGVTISNCKFTHHEKAVLLGASDTHFQDLKMH
VTLAYNIFTNTVHERMPRCRFGFFQIVNNFYDRWDKYAIGGSSNPTILSQGNK
FVAPDFIYKKNVCLRTGAQEPEWMTWNWRTQNDVLENGAIFVASGSDPVLT
AEQNAGMMQAEPGDMVPQLTMNAGVLTCSPGAPC

113477 Amb a 1.3

MGIKQCCYILYFTLALVALLQPVRSAEGVGEILPSVNETRSLQACEALNIIDKC
WRGKADWENNRQALADCAQGFAKGTYGGKWGDVYTVTSNLDDDVANPK
EGTLRFAAAQNRPLWIIFKNDMVINLNQELVVNSDKTIDGRGVKVEIINGGLT
LMNVKNIIIHNINIHDVKVLPGGMIKSNDGPPILRQASDGDTINVAGSSQIWID
HCSLSKSFDGLVDVTLGSTHVTISNCKFTQQSKAILLGADDTHVQDKGMLAT
VAFNMFTDNVDQRMPRCRFGFFQVVNNNYDRWGTYAIGGSSAPTILCQGNR
FLAPDDQIKKNVLARTGTGAAESMAWNWRSDKDLLENGAIFVTSGSDPVLT
PVQSAGMIPAEPGEAAIKLTSSAGVFSCHPGAPC

113476 Amb a 1.2

MGIKHCCYILYFTLALVTLLQPVRSAEDVEEFLPSANETRRSLKACEAHNIIDK
CWRCKADWANNRQALADCAQGFAKGTYGGKHGDVYTVTSDKDDDVANP
KEGTLRFAAAQNRPLWIIFKRNMVIHLNQELVVNSDKTIDGRGVKVNIVNAG
LTLMNVKNIIIHNINIHDIKVCPGGMIKSNDGPPILRQQSDGDAINVAGSSQIWI
DHCSLSKASDGLLDITLGSSHVTVSNCKFTQHQFVLLLGADDTHYQDKGML
ATVAFNMFTDHVDQRMPRCRFGFFQVVNNNYDRWGTYAIGGSSAPTILSQG
NRFFAPDDIIKKNVLARTGTGNAESMSWNWRTDRDLLENGAIFLPSGSDPVL
TPEQKAGMIPAEPGEAVLRLTSSAGVLSCHQGAPC

113475 Amb a 1.1

MGIKHCCYILYFTLALVTLLQPVRSAEDLQEILPVNETRRLTTSGAYNIIDGCW
RGKADWAENRKALADCAQGFGKGTVGGKDGDIYTVTSELDDDVANPKEGT
LRFGAAQNRPLWIIFERDMVIRLDKEMVVNSDKTIDGRGAKVEIINAGFTLNG
VKNVIIHNINMHDVKVNPGGLIKSNDGPAAPRAGSDGDAISISGSSQIWIDHCS
LSKSVDGLVDAKLGTTRLTVSNSLFTQHQFVLLFGAGDENIEDRGMLATVAF
NTFTDNVDQRMPRCRHGFFQVVNNNYDKWGSYAIGGSASPTILSQGNRFCA
PDERSKKNVLGRHGEAAAESMKWNWRTNKDVLENGAIFVASGVDPVLTPE
QSAGMIPAEPGESALSLTSSAGVLSCQPGAPC

Cedar sequences

493634 Cry j IB precursor

MDSPCLVALLVFSFVIGSCFSDNPIDSCWRGDSNWAQNRMKLADCAVGFGS
STMGGKGGDLYTVTNSDDDPVNPPGTLRYGATRDRPLWIIFSGNMNIKLKM
PMYIAGYKTFDGRGAQVYIGNGGPCVFIKRVSNVIIHGLYLYGCSTSVLGNVL
INESFGVEPVHPQDGDALTLRTATNIWIDHNSFSNSSDGLVDVTLTSTGVTISN
NLFFNHHKVMSLGHDDAYSDDKSMKVTVAFNQFGPNCGQRMPRARYGLV
HVANNNYDPWTIYAIGGSSNPTILSEGNSFTAPNESYKKQVTIRIGCKTSSSCS
NWVWQSTQDVFYNGAYFVSSGKYEGGNIYTKKEAFNVENGNATPHLTQNA
GVLTCSLSKRC

493632 Cry j IA precursor

MDSPCLVALLVLSFVIGSCFSDNPIDSCWRGDSNWAQNRMKLADCAVGFGS
STMGGKGGDLYTVTNSDDDPVNPAPGTLRYGATRDRPLWIIFSGNMNIKLK
MPMYIAGYKTFDGRGAQVYIGNGGPCVFIKRVSNVIIHGLHLYGCSTSVLGN
VLINESFGVEPVHPQDGDALTLRTATNIWIDHNSFSNSSDGLVDVTLSSTGVTI
SNNLFFNHHKVMLLGHDDAYSDDKSMKVTVAFNQFGPNCGQRMPRARYGL
VHVANNNYDPWTIYAIGGSSNPTILSEGNSFTAPNESYKKQVTIRIGCKTSSSC
SNWVWQSTQDVFYNGAYFVSSGKYEGGNIYTKKEAFNVENGNATPQLTKN
AGVLTCSLSKRC

1076242 Cry j II precursor - Japanese cedar

MAMKLIAPMAFLAMQLIIMAAAEDQSAQIMLDSVVEKYLRSNRSLRKVEHS
RHDAINIFNVEKYGAVGDGKHDCTEAFSTAWQAACKNPSAMLLVPGSKKFV
VNNLFFNGPCQPHFTFKVDGIIAAYQNPASWKNNRIWLQFAKLTGFTLMGKG
VIDGQGKQWWAGQCKWVNGREICNDRDRPTAIKFDFSTGLIIQGLKLMNSPE
FHLVFGNCEGVKIIGISITAPRDSPNTDGIDIFASKNFHLQKNTIGTGDDCVAIG
TGSSNIVIEDLICGPGHGISIGSLGRENSRAEVSYVHVNGAKFIDTQNGLRIKT
WQGGSGMASHIIYENVEMINSENPILINQFYCTSASACQNQRSAVQIQDVTYK
NIRGTSATAAAIQLKCSDSMPCKDIKLSDISLKLTSGKIASCLNDNANGYFSGH
VIPACKNLSPSAKRKESKSHKHPKTVMVENMRAYDKGNRTRILLGSRPPNCT
NKCHGCSPCKAKLVIVHRIMPQEYYPQRWICSCHGKIYHP

1076241 Cry j II protein - Japanese cedar

MAMKFIAPMAFVAMQLIIMAAAEDQSAQIMLDSDIEQYLRSNRSLRKVEHSR
HDAINIFNVEKYGAVGDGKHDCTEAFSTAWQAACKKPSAMLLVPGNKKFV
VNNLFFNGPCQPHFTFKVDGIIAAYQNPASWKNNRIWLQFAKLTGFTLMGKG
VIDGQGKQWWAGQCKWVNGREICNDRDRPTAIKFDFSTGLIIQGLKLMNSPE
FHLVFGNCEGVKIIGISITAPRDSPNTDGIDIFASKNFHLQKNTIGTGDDCVAIG
TGSSNIVIEDLICGPGHGISIGSLGRENSRAEVSYVHVNGAKFIDTQNGLRIKT
WQGGSGMASHIIYENVEMINSENPILINQFYCTSASACQNQRSAVQIQDVTYK

NIRGTSATAAAIQLKCSDSMPCKDIKLSDISLKLTSGKIASCLNDNANGYFSGH
VIPACKNLSPSAKRKESKSHKHPKTVMVKNMGAYDKGNRTRILLGSRPPNCT
NKCHGCSPCKAKLVIVHRIMPQEYYPQRWMCSRHGKIYHP

541803 Cry j I precursor - Japanese cedar
MDSPCLVALLVLSFVIGSCFSDNPIDSCWRGDSNWAQNRMKLADCAVGFGS
STMGGKGGDLYTVTNSDDDPVNPPGTLRYGATRDRPLWIIFSGNMNIKLKM
PMYIAGYKTFDGRGAQVYIGNGGPCVFIKRVSNVIIHGLHLYGCSTSVLGNVL
INESFGVEPVHPQDGDALTLRTATNIWIDHNSFSNSSDGLVDVTLSSTGVTISN
NLFFNHHKVMLLGHDDAYSDDKSMKVTVAFNQFGPNCGQRMPRARYGLV
HVANNNYDPWTIYAIGGSSNPTILSEGNSFTAPNESYKKQVTIRIGCKTSSSCS
NWVWQSTQDVFYNGAYFVSSGKYEGGNIYTKKEAFNVENGNATPQLTKNA
GVLTCSLSKRC

541802 Cry j I precursor- Japanese cedar
MDSPCLVALLVFSFVIGSCFSDNPIDSCWRGDSNWAQNRMKLADCAVGFGS
STMGGKGGDLYTVTNSDDDPVNPAPGTLRYGATRDRPLWIIFSGNMNIKLK
MPMYIAGYKTFDGRGAQVYIGNGGPCVFIKRVSNVIIHGLYLYGCSTSVLGN
VLINESFGVEPVHPQDGDALTLRTATNIWIDHNSFSNSSDGLVDVTLTSTGVTI
SNNLFFNHHKVMSLGHDDAYSDDKSMKVTVAFNQFGPNCGQRMPRARYGL
VHVANNNYDPWTIYAIGGSSNPTILSEGNSFTAPNESYKKQVTIRIGCKTSSSC
SNWVWQSTQDVFYNGAYFVSSGKYEGGNIYTKKEAFNVENGNATPHLTQN
AGVLTCSLSKRC

Dog

Canis sequences:

Can f 1
MKTLLLTIGFSLIAILQAQDTPALGKDTVAVSGKWYLKAMTADQEVPEKPDS
VTPMILKAQKGGNLEAKITMLTNGQCQNITVVLHKTSEPGKYTAYEGQRVV
FIQPSPVRDHYILYCEGELHGRQIRMAKLLGRDPEQSQEALEDFREFSRAKGL

NQEILELAQSETCSPGGQ

Serum albumin fragment

EAYKSEIAHRYNDLGEEHFRGLVL

Serum albumin fragment

LSSAKERFKCASLQKFGDRAFKAWSVARLSQRFPKADFAEISKVVTDLTKVH
KECCHGDLLECADDRADLAKYMCENQDSISTKLKECCDKPVLEKSQCLAEV
ERDELPGDLPSLAADFVEDKEVCKNYQEAKDVFLGTFLYEYSRRHPEYSVSL
LLRLAKEYEATLEKCCATDDPPTCYAKVLDEFKPLVDEPQNLVKTNCELFEK
LGEYGFQNALLVRYTKKAPQVSTPTLVVEVSRKLGKVGTKCCKKPESERMS
CADDFLS

Can f 2

MQLLLLTVGLALICGLQAQEGNHEEPQGGLEELSGRWHSVALASNKSDLIKP
WGHFRVFIHSMSAKDGNLHGDILIPQDGQCEKVSLTAFKTATSNKFDLEYWG
HNDLYLAEVDPKSYLILYMINQYNDDTSLVAHLMVRDLSRQQDFLPAFESVC
EDIGLHKDQIVVLSDDDRCQGSRD

Additional dog allergen protein (NCBI entrez accession):

1731859

Horse

Equus sequences:

1575778 Equ c1

MKLLLLCLGLILVCAQQEENSDVAIRNFDISKISGEWYSIFLASDVKEKIEENG
SMRVFVDVIRALDNSSLYAEYQTKVNGECTEFPMVFDKTEEDGVYSLNYDG
YNVFRISEFENDEHIILYLVNFDKDRPFQLFEFYAREPDVSPEIKEEFVKIVQKR
GIVKENIIDLTKIDRCFQLRGNGVAQA

3121755 Equ c 2

SQXPQSETDYSQLSGEWNTIYGAASNIXK

Euroglyphus (mite)

Euroglyphus sequences:

Eur m 1 (variant)

TYACSINSVSLPSELDLRSLRTVTPIRMQGGCGSCWAFSGVASTESAYLAYRN
MSLDLAEQELVDCASQNGCHGDTIPRGIEYIQQNGVVQEHYYPYVAREQSC
HRPNAQRYGLKNYCQISPPDSNKIRQALTQTHTAVAVIIGIKDLNAFRHYDGR
TIMQHDNGYQPNYHAVNIVGYGNTQGVDYWIVRNSWDTTWGDNGYGYFA
ANINL

Eur m 1 (variant)

TYACSINSVSLPSELDLRSLRTVTPIRMQGGCGSCWAFSGVASTESAYLAYRN
MSLDLAEQELVDCASQNGCHGDTIPRGIEYIQQNGVVQEHYYPYVAREQSC
HRPNAQRYGLKNYCQISPPDSNKIRQALTQTHTAVAVIIGIKDLNAFRHYDGR
TIMQHDNGYQPNYHAVNIVGYGNTQGVDYWIVRNSWDTTWGDNGYGYFA
ANINL

Eur m 1 (variant)

ETNACSINGNAPAEIDLRQMRTVTPIRMQGGCGSCWAFSGVAATESAYLAY
RNQSLDLAEQELVDCASQHGCHGDTIPRGIEYIQHNGVVQESYYRYVAREQS
CRRPNAQRFGISNYCQIYPPNANKIREALAQTHSAIAVIIGIKDLDAFRHYDGR
TIIQRDNGYQPNYHAVNIVGYSNAQGVDYWIVRNSWDTNWGDNGYGYFAA
NIDL

Eur m 1 (variant)

ETSACRINSVNVPSELDLRSLRTVTPIRMQGGCGSCWAFSGVAATESAYLAY

RNTSLDLSEQELVDCASQHGCHGDTIPRGIEYIQQNGVVEERSYPYVAREQQ
CRRPNSQHYGISNYCQIYPPDVKQIREALTQTHTAIAVIIGIKDLRAFQHYDGR
TIIQHDNGYQPNYHAVNIVGYGSTQGVDYWIVRNSWDTTWGDSGYGYFQA
GNNL

Poa (grass) sequences

113562 POLLEN ALLERGEN POA P 9
MAVQKYTVALFLVALVVGPAASYAADLSYGAPATPAAPAAGYTPAAPAGA
APKATTDEQKMIEKINVGFKAAVAAAGGVPAANKYKTFVATFGAASNKAFA
EALSTEPKGAAVDSSKAALTSKLDAAYKLAYKSAEGATPEAKYDDYVATLS
EALRIIAGTLEVHGVKPAAEEVKATPAGELQVIDKVDAAFKVAATAANAAPA
NDKFTVFEAAFNDAIKASTGGAYQSYKFIPALEAAVKQSYAATVATAPAVK
YTVFETALKKAITAMSQAQKAAKPAAAATGTATAAVGAATGAATAAAGGY
KV

113561 POA P 9
MAVHQYTVALFLAVALVAGPAASYAADVGYGAPATLATPATPAAPAAGYT
PAAPAGAAPKATTDEQKLIEKINAGFKAAVAAAAGVPAVDKYKTFVATFGT
ASNKAFAEALSTEPKGAAAASSNAVLTSKLDAAYKLAYKSAEGATPEAKYD
AYVATLSEALRIIAGTLEVHAVKPAGEEVKAIPAGELQVIDKVDAAFKVAAT
AANAAPANDKFTVFEAAFNDAIKASTGGAYQSYKFIPALEAAVKQSYAATV
ATAPAVKYTVFETALKKAITAMSQAQKAAKPAAAVTATATGAVGAATGAV
GAATGAATAAAGGYKTGAATPTAGGYKV

113560 POA P 9
MDKANGAYKTALKAASAVAPAEKFPVFQATFDKNLKEGLSGPDAVGFAKK
LDAFIQTSYLSTKAAEPKEKFDLFVLSLTEVLRFMAGAVKAPPASKFPAKPAP
KVAAYTPAAPAGAAPKATTDEQKLIEKINVGFKAAVAAAAGVPAASKYKTF
VATFGAASNKAFAEALSTEPKGAAVASSKAVLTSKLDAAYKLAYKSAEGAT
PEAKYDAYVATLSEALRIIAGTLEVHGVKPAAEEVKAIPAGELQVIDKVDAA

FKVAATAANAAPANDKFTVFEAAFNDAIKASTGGAYQSYKFIPALEAAVKQ
SYAATVATAPAVKYTVFETALKKAITAMSQAQKAAKPAAAVTGTATSAVG
AATGAATAAAGGYKV

Cockroach sequences

2833325 Cr p1

MKTALVFAAVVAFVAARFPDHKDYKQLADKQFLAKQRDVLRLFHRVHQHN
ILNDQVEVGIPMTSKQTSATTVPPSGEAVHGVLQEGHARPRGEPFSVNYEKH
REQAIMLYDLLYFANDYDTFYKTACWARDRVNEGMFMYSFSIAVFHRDDM
QGVMLPPPYEVYPYLFVDHDVIHMAQKYWMKNAGSGEHHSHVIPVNFTLR
TQDHLLAYFTSDVNLNAFNTYYRYYYPSWYNTTLYGHNIDRRGEQFYYTYK
QIYARYFLERLSNDLPDVYPFYYSKPVKSAYNPNLRYHNGEEMPVRPSNMY
VTNFDLYYIADIKNYEKRVEDAIDFGYAFDEHMKPHSLYHDVHGMEYLADM
IEGNMDSPNFYFYGSIYHMYHSMIGHIVDPYHKMGLAPSLEHPETVLRDPVF
YQLWKRVDHLFQKYKNRLPRYTHDELAFEGVKVENVDVGKLYTYFEQYD
MSLDMAVYVNNVDQISNVDVQLAVRLNHKPFTYNIEVSSDKAQDVYVAVF
LGPKYDYLGREYDLNDRRHYFVEMDRFPYHVGAGKTVIERNSHDSNIIAPER
DSYRTFYKKVQEAYEGKSQYYVDKGHNYCGYPENLLIPKGKKGGQAYTFY
VIVTPYVKQDEHDFEPYNYKAFSYCGVGSERKYPDNKPLGYPFDRKIYSNDF
YTPNMYFKDVIIFHKKYDEVGVQGH

2231297 Cr p2

INEIHSIIGLPPFVPPSRRHARRGVGINGLIDDVIAILPVDELKALFQEKLETSPD
FKALYDAIRSPEFQSIISTLNAMQRSEHHQNLRDKGVDVDHFIQLIRALFGLSR
AARNLQDDLNDFLHSLEPISPRHRHGLPRQRRRSARVSAYLHADDFHKIITTIE
ALPEFANFYNFLKEHGLDVVDYINEIHSIIGLPPFVPPSRRHARRGVGINGLIDD
VIAILPVDELKALFQEKLETSPDFKALYDAIRSPEFQSIISTLNAMPEYQELLQN
LRDKGVDVDHFIRVDQGTLRTLSSGQRNLQDDLNDFLALIPTDQILAIAMDYL
ANDAEVQELVAYLQSDDFHKIITTIEALPEFANFYNFLKEHGLDVVDYINEIHS
IIGLPPFVPPSQRHARRGVGINGLIDDVIAILPVDELKALFQEKLETSPDFKALY

DAIDLRSSRA

1703445 Bla g 2

MIGLKLVTVLFAVATITHAAELQRVPLYKLVHVFINTQYAGITKIGNQNFLTV
FDSTSCNVVVASQECVGGACVCPNLQKYEKLKPKYISDGNVQVKFFDTGSA
VGRGIEDSLTISNLTTSQQDIVLADELSQEVCILSADVVVGIAAPGCPNALKGK
TVLENFVEENLIAPVFSIHHARFQDGEHFGEIIFGGSDWKYVDGEFTYVPLVG
DDSWKFRLDGVKIGDTTVAPAGTQAIIDTSKAIIVGPKAYVNPINEAIGCVVE
KTTTRRICKLDCSKIPSLPDVTFVINGRNFNISSQYYIQQNGNLCYSGFQPCGH
SDHFFIGDFFVDHYYSEFNWENKTMGFGRSVE
SV

1705483 Bla g 4

AVLALCATDTLANEDCFRHESLVPNLDYERFRGSWIIAAGTSEALTQYKCWI
DRFSYDDALVSKYTDSQGKNRTTIRGRTKFEGNKFTIDYNDKGKAFSAPYSV
LATDYENYAIVEGCPAAANGHVIYVQIRFSVRRFHPKLGDKEMIQHYTLDQV
NQHKKAIEEDLKHFNLKYEDLHSTCH

2326190 Bla g 5

YKLTYCPVKALGEPIRFLLSYGEKDFEDYRFQEGDWPNLKPSMPFGKTPVLEI
DGKQTHQSVAISRYLGKQFGLSGKDDWENLEIDMIVDTISDFRAAIANYHYD
ADENSKQKKWDPLKKETIPYYTKKFDEVVKANGGYLAAGKLTWADFYFVA
ILDYLNHMAKEDLVANQPNLKALREKVLGLPAIKAWVAKRPPTDL

Additional cockroach sequences (NCBI Entrez accession numbers):
2580504; 1580797; 1580794; 1362590; 544619; 544618; 15315104; 1580792; 1166573; 1176397; 21047849.

<u>Allergen (general) sequences:</u>

NCBI accession numbers

2739154; 3719257; 3703107; 3687326; 3643813; 3087805; 1864024; 1493836; 1480457; 25910476; 25910474; 1575778; 763532; 746485; 163827; 163823; 3080761; 163825; 3608493; 3581965; 2253610; 2231297; 21047849; 3409499; 3409498; 3409497; 3409496; 3409495; 3409494; 3409493; 3409492; 3409491; 3409490; 34094104; 3409488; 3409487; 3409486; 3409485; 3409484; 3409483; 3409482; 3409481; 3409480; 3409479; 3409478; 3409477; 3409476; 3409475; 3409474; 3409473; 3409472; 3409471; 3409470; 3409469; 3409468; 3409467; 3409466; 3409465; 3409464; 3409463; 3409462; 3409461; 3409460; 3409459; 3409458; 3409457; 3409456; 3318885; 3396070 ; 3367732; 1916805; 3337403; 2851457; 2851456; 1351295; 549187; 136467; 1173367; 2499810; 2498582; 2498581; 1346478; 1171009; 126608; 114091; 2506771; 1706660; 1169665; 1169531; 232086; 4161048; 114922; 2497701; 1703232; 1703233; 1703233; 1703232; 3287877; 3122132; 3182907; 3121758; 3121756; 3121755; 3121746; 3121745; 3319925; 3319923; 3319921; 3319651; 33187104; 3318779; 3309647; 3309047; 3309045; 3309043; 3309041; 3309039; 3288200; 3288068; 2924494; 3256212; 3256210; 3243234; 3210053; 3210052; 3210051; 3210050; 3210049; 3210048; 3210047; 3210046; 3210045; 3210044; 3210043; 3210042; 3210041; 3210040; 3210039; 3210038; 3210037; 3210036; 3210035; 3210034; 3210033; 3210032; 3210031; 3210030; 3210029; 3210028; 3210027; 3210026; 3210025; 3210024; 3210023; 3210022; 3210021; 3210020; 3210019; 3210018; 3210017; 3210016; 3210015; 3210014; 3210013; 3210012; 3210011; 3210010; 3210009; 3210008; 3210007; 3210006; 3210005; 3210004; 3210003; 3210002; 3210001; 3210000; 3209999; 3201547; 2781152; 2392605; 2392604; 2781014; 1942360; 2554672; 2392209; 3114481; 3114480; 2981657; 3183706; 3152922 ; 3135503 ; 3135501; 3135499; 3135497; 2414158; 1321733; 1321731; 1321728; 1321726; 1321724; 1321722; 1321720; 1321718; 1321716; 1321714; 1321712; 3095075; 3062795; 3062793; 3062791; 2266625; 2266623; 2182106; 3044216; 2154736; 3021324; 3004467; 3005841; 3005839; 3004485; 3004473; 3004471; 3004469; 3004465; 2440053; 1805730; 2970629 ; 29591048; 2935527 ; 2935416; 809536; 730091; 585279; 584968; 2498195; 2833325; 2498604; 2498317; 2498299; 2493414; 2498586; 2498585; 2498576; 2497749; 2493446; 2493445; 1513216 ;

729944; 2498099; 548449; 465054; 465053; 465052; 548671; 548670; 548660; 548658; 548657; 2832430; 232084; 2500822; 2498118; 2498119; 2498119; 2498118; 1708296; 1708793; 416607; 416608; 416608; 416607; 2499791; 2498580; 2498579; 2498578; 2498577; 2497750; 1705483; 1703445; 1709542; 1709545; 17105104; 1352699; 1346568; 1346323; 1346322; 2507248; |1352240; 1352239; 1352237; 1352229; 1351935; 1350779; 1346806; 1346804; 1346803; 1170095; 1168701; 1352506; 1171011; 1171008; 1171005; 1171004; 1171002; 1171001; 1168710; 1168709; 1168708; 1168707; 1168706; 1168705; 1168704; 1168703; 1168702; 1168696; 1168391; 1168390; 1168348; 1173075; 1173074; 1173071; 1169290; 11610470; 1168402; 729764; 729320; 729979; 729970; 729315; 730050; 730049; 730048; 549194; 549193; 549192; 549191; 549190; 5491104; 549188; 549185; 549184; 549183; 549182; 549181; 549180; 549179; 464471; 585290; 416731; 1169666; 113478; 113479; 113477; 113476; 113475; 130975; 119656; 113562; 113561; 113560; 416610; 126387; 126386; 126385; 132270; 416611; 416612; 416612; 416611; 730035; 127205; 1352238; 125887; 549186; 137395; 730036; 133174; 114090; 131112; 126949; 129293; 124757; 129501; 416636; 2801531; 2796177; 2796175; 2677826; 2735118; 2735116; 2735114; 2735112; 2735110; 2735108; 2735106 ; 2735104; 2735102 ; 2735100 ; 2735098 ; 2735096 ; 2707295 ; 2154730; 2154728; 1684720; 2580504 ; 2465137; 2465135; 2465133; 2465131; 2465129; 2465127; 2564228; 2564226; 2564224; 2564222; 2564220; 2051993; 1313972; 1313970; 1313968; 1313966; 2443824 ; 2488684; 2488683; 2488682; 2488681; 2488680; 2488679; 2488678; 2326190 ; 2464905; 2415702; 2415700; 2415698; 2398759; 2398757; 2353266 ; 2338288 ; 1167836; 414703 ; 2276458 ; 1684718 ; 2293571 ; 1580797 ; 1580794 ; 2245508 ; 2245060; 1261972; 2190552 ; 1881574 ; 511953 ; 1532058; 1532056; 1532054; 1359436; 666007; 487661; 217308; 1731859; 217306; 217304; 1545803; 1514943; 577696; 516728; 506858; 493634; 493632; 2154734; 2154732; 543659; 1086046; 1086045; 2147643; 2147642; 1086003; 1086002; 1086001; 543675; 543623; 543509; 543491; 1364099; 2147108; 2147107; 1364001; 1085628; 631913; 631912; 631911; 2147092; 477301; 543482; 345521; 542131; 542130; 542129; 100636; 2146809; 480443; 2114497; 2144915; 72355; 71728; 319828; 1082946; 1082945; 1082944; 539716; 539715;

423193; 423192; 423191; 423190; 1079187; 627190; 6271104; 627188; 627187; 482382; 1362656; 627186; 627185; 627182; 482381; 85299; 85298; 2133756; 2133755; 1079186; 627181; 321044; 321043; 112559; 112558; 1362590; 2133564; 1085122; 10710471; 627144; 627143; 627142; 627141; 280576; 102835; 102834; 102833; 102832; 84703; 84702; 84700; 84699; 84698; 84696; 477888; 477505; 102575; 102572; 478272; 2130094; 629813; 629812; 542172; 542168; 542167; 481432; 320620; 280414; 626029; 542132; 320615; 320614; 100638; 100637; 100635; 82449; 320611; 320610; 280409; 320607; 320606; 539051; 539050; 539049; 539048; 322803; 280407; 100501; 100498; 100497; 100496; 1362137; 1362136; 1362135; 1362134; 1362133; 1362132; 1362131; 1362130; 1362129; 1362128; 100478; 21291041; 1076531; 1362049; 1076486; 2129817; 2129816; 2129815; 2129814; 2129813; 2129812; 2129805; 2129804; 2129802; 2129801; 2129800; 2129799; 479902; 479901; 2129477; 1076247; 629480; 1076242; 1076241; 541803; 541802; 280372; 280371; 1361968; 1361967; 1361966; 1361965; 1361964; 1361963; 1361962; 1361961; 1361960; 1361959; 320546; 2119763; 543622; 541804; 478825; 478824; 478823; 421788; 320545; 81444; 626037; 626028; 539056; 483123; 481398; 481397; 100733; 100732; 100639; 625532; 1083651; 322674; 322673; 81719; 81718; 2118430; 2118429; 2118428; 2118427; 419801; 419800; 419799; 419798; 282991; 100691; 322995; 322994; 101824; 626077; 414553 ; 398830 ; 1311457; 1916292 ; 1911819; 1911818; 1911659; 1911582; 467629; 467627; 467619 ; 467617 ; 915347; 1871507; 1322185; 1322183; 1047645 ; 1047647 ; 1850544 ; 1850542 ; 1850540 ; 2810417; 452742; 1842045 ; 1839305; 1836011; 1836010; 1829900; 18291049; 18291048; 18291047; 18291046; 18291045; 18291044; 1825459 ; 18010487 ; 159653 ; 1773369 ; 1769849; 1769847; 608690 ; 1040877 ; 1040875; 1438761; 1311513; 1311512; 1311511; 1311510; 1311509; 13116104; 1246120; 1246119; 1246118; 1246117; 1246116; 1478293; 1478292; 1311642; 1174278; 1174276; 1086972; 1086974; 1086976; 1086978; 1086978; 1086976; 1086974; 1086972; 999009; 999356; 999355; 994866; 994865; 913758; 913757; 913756; 913285; 913283; 926885; 807138; 632782; 601807; 546852; 633938; 544619; 544618; 453094; 451275; 451274; 407610; 407609; 404371; 409328; 299551; 299550; 264742; 261407; 255657; 250902; 250525;

1613674; 1613673; 1613672; 1613671; 1613670; 1613304; 1613303; 1613302; 1613240; 1613239; 1613238; 1612181; 1612180; 1612179; 1612178; 1612177; 1612176; 1612175; 1612174; 1612173; 1612172; 1612171; 1612170; 1612169; 1612168; 1612167; 1612166; 1612165; 1612164; 1612163; 1612162; 1612161; 1612160; 1612159; 1612158; 1612157; 1612156; 1612155; 1612154; 1612153; 1612152; 1612151; 1612150; 1612149; 1612148; 1612147; 1612146; 1612145; 1612144; 1612143; 1612142; 1612141; 1612140; 1612139; 1093120; 447712; 447711; 447710; 1587177; 158542; 1582223; 1582222; 15315104 ; 1580792 ; 886215; 15451047; 15451045; 15451043; 15451041; 15458104; 1545887; 1545885; 1545883; 1545881; 1545879; 1545877; 1545875; 166486 ; 1498496 ; 1460058; 972513; 1009442 ; 1009440 ; 1009438 ; 1009436 ; 1009434 ; 7413 ; 1421808 ; 551228; 452606 ; 32905; 1377859 ; 1364213; 1364212; 395407; 22690 ; 22688 ; 22686; 22684 ; 488605 ; 17680 ; 1052817 ; 1008445 ; 1008443 ; 992612; 706811 ; 886683; 747852 ; 939932 ; 19003 ; 1247377 ; 1247375; 1247373; 862307 ; 312284 ; 999462; 999460 ; 999458 ; 587450 ; 763064 ; 886209 ; 1176397 ; 1173557 ; 902012; 997915; 997914; 997913; 997912; 997911; 997910; 99790; 997908; 997907; 997906; 997905; 997904; 997903; 997902; 997901; 997900; 9971049; 9971048; 9971047; 9971046; 9971045; 9971044; 9971043; 9971042; 910984; 910983; 910982; 910981; 511604 ; 169631 ; 169629 ; 169627 ; 168316 ; 168314 ; 607633 ; 555616; 293902 ; 485371 ; 455288 ; 166447 ; 166445 ; 166443 ; 166435 ; 162551 ; 160780; 552080 ; 156719 ; 156715 ; 515957 ; 515956 ; 515955 ; 515954 ; 515953 ; 459163; 166953 ; 386678 ; 169865.

Particularly preferred allergens /antigens include: cat dander protein Fel d1; House dust mite proteins Der P1, Der P2 and Der P7; Ragweed protein amb a 1.1, a 1.2, a1.3 or a1.4; Rye grass proteins lol p1 and lol p5; Timothy grass proteins phl p1 and phl p5; Bermuda grass protein Cyn d 5; Alternaria alternate proteins Alt a 1, Alt a 2 and Enolase (Alt a 6); Birch protein Bet v1 and P14; German Cockroach proteins Bla g 1, Bla g 2, Bla g 3, Bla g 4, Bla g 5 and Bla g 6; Mugwort protein Art v 1; Russian thistle protein Sal k 1 and Sal k 2; peanut Ara h1, Ara h2, Ara h3, Ara h4, Ara h5, Ara h6, plant profilins or lipid transfer proteins or a human leukocyte antigen.

*Delivery methods*

Once formulated the compositions of the invention can be delivered to a subject *in vivo* using a variety of known routes and techniques. For example, a composition can be provided as an injectable solution, suspension or emulsion and administered via parenteral, subcutaneous, epidermal, intradermal, intramuscular, intraarterial, intraperitoneal, intravenous injection using a conventional needle and syringe, or using a liquid jet injection system. Compositions can also be administered topically to skin or mucosal tissue, such as nasally, intratracheally, intestinal, rectally or vaginally, or provided as a finely divided spray suitable for respiratory or pulmonary administration. Other modes of administration include oral administration, suppositories, sublingual administration, and active or passive transdermal delivery techniques.

Where a peptide of the invention is to be administered, it is preferred to administer the peptide to a site in the body where it will have the ability to contact suitable antigen presenting cells, and where it, or they, will have the opportunity to contact T cells of the individual. Where an APC is to be administered, it is preferred to administer the APC to a site in the body where it will have the ability to contact, and activate, suitable T cells of the individual.

*Delivery regimes*

Administration of the peptides/polynucleotides/cells (such as the composition containing a plurality of peptides) may be by any suitable method as described above. Suitable amounts of the peptide may be determined empirically, but typically are in the range given below. A single administration of each peptide may be sufficient to have a beneficial effect for the patient, but it will be appreciated that it may be beneficial if the peptide is administered more than once, in which case typical administration regimes may be, for example, once or twice a week for 2-4 weeks every 6 months, or once a day for a week every four to six months. As will be appreciated, each peptide or polynucleotide, or combination of peptides and/or polynucleotides may be administered to a patient singly or in combination.

Dosages for administration will depend upon a number of factors including the nature of the composition, the route of administration and the schedule and timing of the administration regime. Suitable doses of a molecule of the invention may be in the order of up to 15μg, up to 20μg, up to 25μg, up to 30μg, up to 50μg, up to 100μg, up to 500 μg or more per administration. Suitable doses may be less than 15μg, but at least 1ng, or at least 2ng, or at least 5ng, or at least 50ng, or least 100ng, or at least 500ng, or at least 1μg, or at least 10μg. For some molecules of the invention, the dose used may be higher, for example, up to 1 mg, up to 2 mg, up to 3 mg, up to 4 mg, up to 5 mg or higher. Such doses may be provided in a liquid formulation, at a concentration suitable to allow an appropriate volume for administration by the selected route.

*Kits*

The invention also relates to a combination of components described herein suitable for use in a treatment of the invention which are packaged in the form of a kit in a container. Such kits may comprise a series of components to allow for a treatment of the invention. For example, a kit may comprise one or more different peptides, polynucleotides and/or cells of the invention, or one or more peptides, polynucleotides or cells of the invention and one or more additional therapeutic agents suitable for simultaneous administration, or for sequential or separate administration. The kit may optionally contain other suitable reagent(s) or instructions and the like.

The invention is illustrated by the following Examples:

Example 1

*MHC Class II binding search*

The aim of this study is to identify a distinct panel of peptides with strong affinities for the seven most common human MHC Class II HLA-DRB1* allotypes (covering in total around 63% of the allotypes found in the average Caucasian population). In order to identify binding peptides in the House Dust Mite (HDM) allergens, Der p 1, Der p 2 and Der p 7, in vitro binding assays have been performed on a subset of

peptides from these allergenic proteins. Peptides for testing in the binding assays were initially identified by an *in silico* approach known as "peptide threading" (carried out by Biovation, Ltd.,Aberdeen, Scotland, UK). This is a bioinformatic analysis of consecutive peptides from a sequence for the potential to be accommodated within the binding groove of MHC class II HLA-DR molecules. This subset of peptides was pre-screened for solubility in an aqueous, acidic milieu and a final panel of 44 peptides selected for testing in an in vitro MHC Class II binding assay.

*Methods*

The assay employed is a competitive MHC class II binding assay, wherein each peptide is analysed for its ability to displace a known control binder from each of the human MHC class II allotypes investigated. The allotypes and control peptides used in this study are shown in Table 2:

| Allotype | Control Peptide | Sequence |
|---|---|---|
| DRB1*0101 | Influenza haemagglutinin 307-319 | PKYVKQNTLKLAT |
| DRB1*0301 | Myco. tuberculosis/leprae hsp 65 2-16 | AKTIAYDEEARRGLE |
| DRB1*0401 | Influenza haemagglutinin 307-319 | PKYVKQNTLKLAT |
| DRB1*0701 | Influenza haemagglutinin 307-319 | PKYVKQNTLKLAT |
| DRB1*1101 | Influenza haemagglutinin 307-319 | PKYVKQNTLKLAT |
| DRB1*1301 | HLA-DQB1*0603 21-36 | TERVRLVTRHIYNREE |
| DRB1*1501 | Human myelin basic protein 85-99 | ENPVVHFFKNIVTPR |
| DQB1*0602 | Human Insulin B 1-15 | FVNQHLCGSHLVEAL |

Table 2. Control peptides used in the in vitro binding assays

Each of the 44 HDM peptides (which are shown in Tables 3A and 3B) were analysed in the competition assay and screened for relative binding compared to the control peptide. Due to the nature of the competitive assay the data for each peptide is represented as a ratio of its own IC50 to that of the control peptide. Thus, a peptide that has an IC50 value that is parity to the control peptide has an identical binding

affinity, while peptides with a ratio less than one have a higher affinity and those with a ratio greater than one have a lower affinity.

*Results*

Solubility in aqueous solution is an essential criterion for a peptide to be an effective therapeutic agent. Therefore, as a consequence of the solubility screen we will have eliminated very hydrophobic peptides with a high frequency of large hydrophobic amino acid residues in multiple binding registers. This is a characteristic of promiscuous HLA-DRB1* binders. The data from the binding assays is shown in Table 3B. The relative binding of each peptide is shown for each of the allotypes in the study. The data shows that 24 of the 44 peptides tested bound to one or more of the MHC Class II allotypes. A range of cross-reactivity is seen with 5 peptides binding only one allotype, 8 peptides binding two, 9 peptides binding three and two peptides binding four different MHC Class II allotypes (red). It would also be expected that such peptides would have the ability to bind similar allotypes that have not been tested through the homology of MHC structures. This can be seen in the cross-reactivity of peptides for DRB1*0101, *0401, *0701 and *1101 in several cases here. Also shown is the solubility status of the peptide at the highest concentrations in the aqueous solution of the binding assay. The value illustrates the lowest concentration at which an insoluble white precipitate is seen. There appears to be no significant nonspecific effect of the formation of precipitate in the assays. Several peptides that precipitate at high concentrations also bind to MHC class II; however, several also show no ability to compete with the control peptides. It is to be expected that peptides liable to form precipitates may exhibit high affinity and promiscuous binding due to the presence of many hydrophobic residues.

The % purity of the peptides is indicated in Table 3A. This is of significance as purities were seen to vary from 60-90%. This would have a considerable effect on the ability of a peptide to compete if it is relatively impure. For example, HDM23A and HDM32 show low affinity binding; however, they are of reduced purity (66.7% and 68.7% respectively) compared to other HDM peptides. Therefore, if purity is taken into consideration, they may in fact have an equivalent affinity to a peptide of a

higher purity.

It can be seen that some MHC Class II allotypes bind to more peptides than others; this is probably to be expected as there is variability between the pocket positions in the different MHC class II binding grooves. There are however, also a number of well-characterised differences between the affinities of the various control peptides. Clearly a high affinity control peptide will be more difficult to displace by the competing HDM peptide resulting in the identification of fewer binding peptides. This can be illustrated by the data presented here. For example, the Influenza Haemagglutinin 307-319 control peptide, has varying affinity according to the allotype, where DRB1*0101 >*0401>*0701>*1101. This is reflected in the number of binders to each of the allotypes, where DRB1*0101 has the lowest number of binders (5) and DRB1*1101 has the highest(14). Furthermore, the binding assay for DRB1*1501 is very stringent due to the high affinity of Myelin Basic Protein 85-99 for this allotype. In the high stringency screen the Fel d 1 peptide EQVAQYKALPVVLENA, that was tested in an earlier study, gave a ratio of 0.97 indicating that high affinity binders could be identified at this stringency.

In addition, to identify lower affinity binders, the assay was also carried out under less stringent conditions. All the Der p binding peptides were seen to have a high ratio when tested against this allotype, showing they were low affinity binders compared to the control peptide. The DQA1*0102/DQB1*0602 binding assay uses a peptide from the B-chain of human insulin which is of lower affinity compared to those used in the DR assays. This dictates that the DQ assay is very sensitive and tends to produce very low ratio values for the strongest binders to this MHC Class II allotype. This sensitivity also accounts for the relatively higher number of DQ binding peptides within the panel screened. Finally, on closer analysis, the peptides identified as ligands for the DRB1*0101,*0401, *0701 superfamily, are found to incorporate a motif that is characteristic of promiscuous binders to this family of allotypes where: P1 = Y, F, W, L, I, V, or M (Large aromatic or hydrophobic residue), P6 = S, T, C, A, P, V, I, M (small, non-charged residue)
Out of the 16 peptides (e.g. HDM 21B RGKPFQLEAVFEANQNT) identified as binders to all or a combination of these 3 allotypes, 14 (87.5%) contain this motif,

which suggests that these are promiscuous binders with a range of affinities for the 1-4-7 allotypes.

*Conclusions*

A range of peptides have been shown to have the capacity to bind the MHC Class II allotypes tested and can be tested for their ability to stimulate in vitro proliferation of CD4+ T cells and to stimulate T cell cytokine secretion.

Table 3A

| Peptide | Sequence | Residues in parent | % purity | Solubility test | Precipitation in assay |
|---|---|---|---|---|---|
| HDM01 | IDLRQMRTVTPIR | 112-124 | 79.2 | YES | None |
| HDM02 | RTVTPIRMQGGCG | 118-130 | 79.6 | YES | None |
| HDM03C | RNQSLDLAEQELVDCASQH | 149-167 | 60.1 | YES | None |
| HDM05 | EYIQHNGVVQESY | 179-191 | 77.5 | YES | None |
| HDM06 | RYVAREQSCRRPN | 193-205 | 79.7 | YES | None |
| HDM07 | PNVNKIREALAQT | 220-232 | 88.6 | YES | None |
| HDM08 | NKIREALAQTHSA | 223-235 | 87.6 | YES | None |
| HDM09A | REALAQTHSAIAVI | 226-239 | 69.6 | YES | 1000μM (2.9mg/ml) |
| HDM11 | IGIKDLDAFRHYD | 240-252 | 77.6 | YES | None |
| HDM12 | KDLDAFRHYDGRT | 243-255 | 72.9 | YES | None |
| HDM13 | RTIIQRDNGYQPNY | 254-267 | 70.7 | NO | None |
| HDM16A | RNSWDTNWGDNGYG | 287-300 | 70.00 | YES | None |
| HDM17 | NSVNVPSELDLRSLRT | 105-120 | 74.5 | YES | None |
| HDM19 | DQVDVKDCANHEIKK | 18-32 | 81.4 | YES | None |
| HDM20 | CIIHRGKPFQLEA | 44-56 | 77.4 | YES | None |
| HDM21 | KPFQLEAVFEANQNT | 50-64 | 88.7 | YES | 200μM (0.3 mg/ml) |
| HDM21A | KPFQLEAVFEANQNTK | 50-65 | 90.10 | YES | 5000μM (9.3 mg/ml) |
| HDM21B | RGKPFQLEAVFEANQNT | 48-64 | 82.60 | YES | 1000μM (1.98mg/ml) |
| HDM22A | EAVFEANQNTKTAK | 55-68 | 90.30 | YES | None |
| HDM23A | DGLEVDVPGIDPNACH | 76-88 | 66.7 | YES | None |
| HDM26A | DGVLACAIATHAKIR | 131-145 | | | 1000μM (1.5mg/ml) |

| | | | | | |
|---|---|---|---|---|---|
| HDM27 | AKIEIKASLDGLE | 67-79 | 65.9 | YES | 1000μM (1.4mg/ml) |
| HDM28 | KAVDEAVAAIEKS | 31-43 | 86.8 | YES | 1000μM (1.3mg/ml) |
| HDM29 | ETFDPMKVPDHSD | 44-56 | 84.7 | YES | None |
| HDM29A | ETFDPMKVPDHSDK | 44-57 | 91.7 | YES | None |
| HDM29B | KSETFDPMKVPDHSD | 42-56 | 92.5 | YES | 1000μM (1.7mg/ml) |
| HDM30 | DKFERHIGIIDLK | 56-68 | 81.4 | YES | 5000μM (7.9mg/ml) |
| HDM31 | IGIIDLKGELDMRN | 62-75 | | | 1000μM (1.8mg/ml) |
| HDM31A | HIGIIDLKGELDMRN | 61-75 | 66.40 | YES | 1000μM (1.7mg/ml) |
| HDM32 | IDLKGELDMRNIQ | 65-77 | 68.7 | YES | 5000μM (7.7mg/ml) |
| HDM32A | IDLKGELDMRNIQVR | 65-79 | 85.20 | YES | 5000μM (9.0mg/ml) |
| HDM33 | LDMRNIQVRGLKQ | 71-83 | 70.3 | YES | None |
| HDM34 | RNIQVRGLKQMKRVG | 74-88 | 74.7 | YES | None |
| HDM35 | RGLKQMKRVGDAN | 79-91 | 84.00 | YES | None |
| HDM36 | KRVGDANVKSEDG | 85-97 | 82.9 | YES | None |
| HDM37 | ANVKSEDGVVKAH | 90-102 | 76.5 | YES | None |
| HDM39 | DDVVSMEYDLAYK | 109-121 | 84.9 | NO* | None |
| HDM39A | HDDVVSMEYDLAYKL | 108-121 | 80.9 | YES | 1000μM (1.8mg/ml) |
| HDM40A | VSMEYDLAYKLGDLH | 112-124 | 66.9 | YES | 1000μM (1.8mg/ml) |
| HDM48 | TAIFQDTVRAEMTK | 187-200 | 79.1 | YES | 1000μM (1.6mg/ml) |
| HDM49 | DTVRAEMTKVLAP | 192-204 | 69.5 | YES | None |

| | | | | | |
|---|---|---|---|---|---|
| HDM50 | KVLAPAFKKELE R | 200-212 | 90.8 | YES | None |
| HDM51 | VDFKGELAMRNI E | 65-77 | 79.8 | YES | 1000μM (1.5mg/ml) |
| HDM51A | VDFKGELAMRNI EAR | 65-79 | 82.1 | YES | None |

Table 3B

| Peptide | DRB1*0101 | DRB1*0301 | DRB1*0401 | DRB1*0701 | DRB1*1101 | DRB1*1301 | DRB1*1501 | DQA1*0102 DQB1*0602 |
|---|---|---|---|---|---|---|---|---|
| HDM01 | | 19.23 | | 16 | | | | |
| HDM02 | | | | | | | 80 | 0.03 |
| HDM03C | | | | | | | | 0.16 |
| HDM05 | | | | | | | | |
| HDM06 | | | | | 30.36 | | | 0.86 |
| HDM07 | | | | | | | | |
| HDM08 | | | | | | | | |
| HDM09A | | | | 0.49 | 21.15 | | 200 | |
| HDM11 | | | | | | | | |
| HDM12 | | | | | | | | |
| HDM13 | | | | | | | | |
| HDM16A | | | | | | | | |
| HDM17 | | | | | | | | |
| HDM19 | | | | | | | | |
| HDM20 | | | | 1.1 | 28 | | 242.11 | 2.37 |
| HDM21 | 92 | | 11.15 | | 11.73 | | | |
| HDM21A | 200 | | 52.17 | | 10.27 | | | |
| HDM21B | 13.5 | | 0.78 | | 4.1 | | | |
| HDM22A | | | 328.6 | | 80 | | | |
| HDM23A | | 347 | | | | | | 0.76 |
| HDM26A | 42.3 | | | 16.28 | | | | 0.61 |
| HDM27 | | | | | | | | |
| HDM28 | | | | | | | | |
| HDM29 | | | | | | | | |
| HDM29A | | | | | | | | |
| HDM29B | | | | | | | | |
| HDM30 | | | | 6.2 | | | | |
| HDM31 | | | | | | | | |
| HDM31A | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| HDM32A | | | | | | | | |
| HDM33 | | | | 46.51 | 41.5 | 263.16 | | |
| HDM34 | | | | | 3.38 | 3.7 | 769.23 | |
| HDM35 | | | | | 1.26 | | | |
| HDM36 | | | | | | | | |
| HDM37 | | | | | | | | |
| HDM39 | | | | | | | | |
| HDM39A | | | 76.19 | 0.71 | | | | 0.1 |
| HDM40A | | | | 2.29 | 6 | | | |
| HDM48 | | 211.26 | | 15.71 | 13.57 | | | |
| HDM49 | | | | | | | 671.43 | 1.7 |
| HDM50 | | | | | | | | |
| HDM51 | | | | 20.93 | 30.91 | | | |

Example 2

*Homology search*

The sequences of each of the 24 peptides identified above as MHC Class II-binding were used to probe the sequence of the alternative protein in the dust mite allergen group from which the parent sequence derived. For example, peptide HDM01 in Table 3A is from Der p 1, therefore the sequence of HDM01 was used to probe for a homologous sequence in Der f 1. The same practice was applied for all 24 peptides identified above. The peptides identified in Example 1 and Example 2 are shown in Tables 4 to 6.

Table 4

| Peptide in Table 3A/B | Parent molecule | Sequence | Residues in parent | SEQ ID NO: |
|---|---|---|---|---|
| HDM01 | *Der p 1* | IDLRQMRTVTPIR | 112-124 | 1 |
| | *Der f 1* | *LDLRSLRTVTPIR* | *113-125* | *25* |
| | | | | |
| HDM02 | *Der p 1* | RTVTPIRMQGGCG | 118-130 | 2 |
| | *Der f 1* | *RTVTPIRMQGGCG* | *119-131* | *26* |
| | | | | |
| HDM03C | *Der p 1* | RNQSLDLAEQELVDCASQH | 149-167 | 3 |
| | *Der f 1* | *RNTSLDLSEQELVDCASQH* | *150-168* | *27* |
| | | | | |
| HDM06 | *Der p 1* | RYVAREQSCRRPN | 193-205 | 4 |
| | *Der f 1* | *PYVAREQRCRRPN* | *194-206* | *28* |
| | | | | |
| HDM09A | *Der p 1* | REALAQTHSAIAVI | 226-239 | 5 |
| | *Der f 1* | *REALTQTHTAIAVI* | *227-240* | *29* |

Table 5

| Peptide in Table 3A/B | Parent molecule | Sequence | Residues in parent | SEQ ID NO: |
|---|---|---|---|---|
| HDM19 | *Der p 2* | DQVDVKDCANHEIKK | 18-32 | 6 |
| | *Der f 2* | DQVDVKDCANNEIKK | *18-32* | *30* |
| | | | | |
| HDM20 | *Der p 2* | CIIHRGKPFQLEA | 44-56 | 7 |
| | *Der f 2* | CIIHRGKPFTLEA | *44-56* | *31* |
| | | | | |
| HDM21 | *Der p 2* | KPFQLEAVFEANQNT | 50-64 | 8 |
| | *Der f 2* | *KPFTLEALFDANQNT* | *50-64* | *32* |
| | | | | |
| HDM21A | *Der p 2* | KPFQLEAVFEANQNTK | 50-65 | 9 |
| | *Der f 2* | *KPFTLEALFDANQNTK* | *50-65* | *33* |
| | | | | |
| HDM21B | *Der p 2* | RGKPFQLEAVFEANQNT | 48-64 | 10 |
| | *Der f 2* | *RGKPFTLEALFDANQNT* | *48-64* | *34* |
| | | | | |

| | | | | |
|---|---|---|---|---|
| HDM22A | *Der p 2* | EAVFEANQNTKTAK | 55-68 | 11 |
| | *Der f 2* | *EALFDANQNTKTAK* | *55-68* | *35* |
| | | | | |
| HDM23A | *Der p 2* | DGLEVDVPGIDPNACH | 76-88 | 12 |
| | *Der f 2* | *DGLEIDVPGIDTNACH* | *76-88* | *36* |
| | | | | |
| HDM26A | *Der p 2* | DGVLACAIATHAKIR | 131-145 | 13 |
| | *Der f 2* | *NGVLACAIATHGKIR* | *131-145* | *37* |

Table 6

| Peptide in Table 3A/B | Parent molecule | Sequence | Residues in parent | SEQ ID NO: |
|---|---|---|---|---|
| HDM30 | *Der p 7* | DKFERHIGIIDLK | 56-68 | 14 |
| | *Der f 7* | *DKFERHVGIVDFK* | *56-68* | *38* |
| | | | | |
| HDM32 | *Der p 7* | IDLKGELDMRNIQ | 65-77 | 15 |
| | *Der f 7* | *VDFKGELAMRNIE* | *65-77* | *39* |
| | | | | |
| HDM33 | *Der p 7* | LDMRNIQVRGLKQ | 71-83 | 16 |
| | *Der f 7* | *LAMRNIEARGLKQ* | *71-83* | *40* |
| | | | | |
| HDM34 | *Der p 7* | RNIQVRGLKQMKRVG | 74-88 | 17 |
| | *Der f 7* | *RNIEARGLKQMKRQG* | *74-88* | *41* |
| | | | | |
| HDM35 | *Der p 7* | RGLKQMKRVGDAN | 79-91 | 18 |
| | *Der f 7* | *RGLKQMKRQGDAN* | *79-91* | *42* |
| | | | | |
| HDM39A | *Der p 7* | HDDVVSMEYDLAYKL | 108-122 | 19 |
| | *Der f 7* | *HDDIVSMEYDLAYKL* | *108-122* | *43* |
| | | | | |
| HDM40A | *Der p 7* | VSMEYDLAYKLGDLH | 112-126 | 20 |
| | *Der f 7* | *VSMEYDLAYKLGDLH* | *112-126* | *44* |
| | | | | |
| HDM48 | *Der p 7* | TAIFQDTVRAEMTK | 187-200 | 21 |
| | *Der f 7* | *TAIFQDTVRKEMTK* | *187-200* | *45* |
| | | | | |
| HDM49 | *Der p 7* | DTVRAEMTKVLAP | 192-204 | 22 |
| | *Der f 7* | *DTVRKEMTKVLAP* | *192-204* | *46* |
| | | | | |
| HDM51 | *Der f 7* | VDFKGELAMRNIE | 65-77 | 23 |
| | *Der p 7* | *IDLKGELDMRNIQ* | *65-77* | *15* |
| | | | | |
| HDM51A | *Der f 7* | VDFKGELAMRNIEAR | 65-79 | 24 |
| | *Der p 7* | *IDLKGELDMRNIQVR* | *65-79* | *47* |

In Table 4, the sequence of Der p 1 from which the "residues in parent" positions are derived is the publically available sequence with NCBI Accession No. P08176. The corresponding sequences for Der p 2 (Table 5) and Der p 7 (Table 6) are NCBI Accession Nos. P49278 and P49273, respectively. The sequence for Der f 1 is taken from NCBI Accession No. P16311, Der f 2 is from NCBI Accession No. Q00855 and Der f 7 is from NCBI Accession No. Q26456.

Example 3

*MHC Class II binding search*

The aim of this study is to identify a distinct panel of peptides with strong affinities for the seven most common human MHC Class II HLA-DRB1* allotypes (covering in total around 63% of the allotypes found in the average Caucasian population). In order to identify binding peptides in the major House dust mite allergens Der p 1, Der p 2 and Der p 7. Peptides were identified by an *in silico* approach known as "peptide threading" using the commercially available EpiMatrix algorithm (EpiVax Inc.) This is a bioinformatic analysis of peptides from a sequence for the potential to be accommodated within the binding groove of MHC class II HLA-DR molecules. EpiMatrix is a matrix-based algorithm that ranks 10 amino acid long segments, overlapping by 9 amino acids, from any polypeptide sequence by estimated probability of binding to each of the selected MHC molecules. (De Groot et al., AIDS Research and Human Retroviruses 13:539-41 (1997)). The procedure for developing matrix motifs was published by Schafer et al, 16 Vaccine 1998 (1998). In this Example, binding potential for HLA DR1, DR2, DR3, DR4, DR7, DR8, DR11, DR13 and DR15 is assessed. Putative MHC ligands are selected by scoring each 10-mer frame in a protein sequence. This score is derived by comparing the sequence of the 10-mer to the matrix of 10 amino acid sequences known to bind to each MHC allele. Retrospective studies have demonstrated that EpiMatrix accurately predicts published MHC ligands (Jesdale et al., in Vaccines '97 (Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1997)). Successful Prediction of peptides which bind to multiple MHC molecules has also been confirmed.

Estimated probability of binding to a selected MHC molecule is calculated by EpiMatrix as follows. The peptides are scored by estimating the relative promotion or inhibition of binding for each amino acid, compared to known MHC binders for a given MHC allele. This information is summed across the peptide and a summary score (EMX score) is assigned to the entire peptide. After comparing the EMX score to the scores of known MHC ligands, EpiMatrix arrives at an "estimated binding probability" (abbreviated as EBP, but not strictly a probability). The EBP describes the proportion of peptides with EpiMatrix scores as high or higher that will bind to a given MHC molecule. EBPs range from 100% (highly likely to bind) to less than 1%

(very unlikely to bind).

EpiMatrix analyses were performed on the entire sequence of the Der p 1 as published in the NCBI database (NCBI accession no: P08176). This analysis identified core peptides (and their flanking sequences) derived from the above sequences which are predicted to have good MHC class-II binding. These sequences are shown below in Table 7A. Tables 7B and 7C show the sequences for the equivalent analyses of known sequences of Der p 2 and Der p 7, respectively (NCBI accession nos. P49278 and P49273).

In Tables 7A - C: "Residues in sequence" gives the location of the peptide within the sequences that were analysed. The core peptide (middle amino acids in bold) defines the actual binding sequence that was identified during the analysis. The stabilizing flanks (N-terminal and C-terminal, not bold) were included for use with the core sequence and are typically required to aid manufacture of the peptides. "Number of hits" refers to the number of high predicted binding affinities for all MHC types tested within the sequence. The "EpiMatrix Cluster Score" is derived from the number of hits normalized for the length of the cluster. Cluster Score is thus the excess or shortfall in predicted aggregate MHC binding properties relative to a random peptide standard. A score of 10 or above is considered to indicate broad MHC binding properties. Epivax also analysed hydrophobicity of peptides containing epitopes. Scores of greater than 1 are considered to be unsuitable for administration and/or manufacture.

**TABLE 7A – Der p 1**

| INPUT SEQUENCE (NCBI no.) | RESIDUES IN SEQUENCE (Incl. FLANKS) | SEQUENCE | Hydro-phobicity | EpiMatrix HITS (Excl FLANKS) | EpiMatrix CLUSTER SCORE (Excl FLANKS) | SEQ ID NO: |
|---|---|---|---|---|---|---|
| P08176 | 1 - 21 | **MKIVLAIASLLALSAVY** ARPS | 1.42 | 22 | 38.91 | 105 |
| P08176 | 51 - 67 | LES**VKYVQSNGGAI**NHL | -0.15 | 6 | 10.87 | 106 |
| P08176 | 72 - 88 | LDE**FKNRFLMSAEA**FEH | -0.49 | 6 | 10.55 | 107 |
| P08176 | 111 - 134 | EID**LRQMRTVTPIRMQG GCGS**CWA | -0.24 | 16 | 26.34 | 108 |
| P08176 | 142 - 159 | ESA**YLAYRNQSLDLAEQ** E | -0.91 | 10 | 16.43 | 109 |
| P08176 | 188 - 209 | QES**YYRYVAREQSCRRP NA**QRF | -1.70 | 14 | 24.92 | 110 |

| | | | | | | |
|---|---|---|---|---|---|---|
| P08176 | 296 - 313 | DNGYGYFAANIDLMMIE E | -0.08 | 7 | 10.24 | 111 |

**TABLE 7B – Der p 2**

| INPUT SEQUENCE (NCBI no.) | RESIDUES IN SEQUENCE (Incl. FLANKS) | SEQUENCE | Hydro-phobicity | EpiMatrix HITS (Excl FLANKS) | EpiMatrix CLUSTER SCORE (Excl FLANKS) | SEQ ID NO: |
|---|---|---|---|---|---|---|
| P49278 | 1 - 22 | MMYKILCLSLLVAAVAR DQVDV | 1.24 | 14 | 21.8 | 112 |
| P49278 | 42 - 63 | EPCIIHRGKPFQLEAVF EANQN | -0.50 | 10 | 14.62 | 113 |

**TABLE 7C – Der p 7**

| INPUT SEQUENCE (NCBI no.) | RESIDUES IN SEQUENCE (Incl. FLANKS) | SEQUENCE | Hydro-phobicity | EpiMatrix HITS (Excl FLANKS) | EpiMatrix CLUSTER SCORE (Excl FLANKS) | SEQ ID NO: |
|---|---|---|---|---|---|---|
| P49273 | 1 - 17 | MMKLLLIAAAAFVAVSA | 2.2 | 12 | 20.16 | 114 |
| P49273 | 70 - 92 | ELDMRNIQVRGLKQMKR VGDANV | -0.71 | 9 | 12.3 | 115 |

## Example 4

*Selection of peptides for further testing*

Based on the peptides and epitopes identified in Examples 1 to 3, the inventors selected the peptides shown in Table 8 for further testing. Some of the peptides selected can be considered to be variants of the peptides of Example 1 to 3, but are also considered to be peptides of the invention. In particular, residues in bold in Table 8 indicate alterations from the corresponding residue in the native sequence of the parent protein. These alterations reduce the formation of peptide dimers and improve solubility without diminishing the functionality of a peptide as a T cell epitope. The alterations shown are the replacement of a cysteine (C) in the native sequence with a serine (S) or 5-aminobutyric acid (Б), or cystine (Ĉ) as indicated.

Additionally, some sequences may comprise more or fewer of the residues of the parent protein from which they derive, when compared to the sequences of the peptides of Examples 1 to 3. Thus, such sequences can be considered to represent truncation or extension variants of the peptides of Examples 1 to 3. For example,

Peptide HDM03F corresponds to resides 149-165 of Der p1. HDM03E corresponds to residues 149-167. Accordingly, HDM03F can be considered to be a truncation variant of HDM03E formed by removal of 2 residues from the N terminus of HDM03E. The "residues in parent" positions in Table 8 refer to the sequences of Der p 1, Der p 2 and Der p 7 as used in Tables 4 to 7. Those peptides indicated in Table 8 which have an N terminal glutamate (E) residue, for example HDM03K, L, V and W, may have the glutamate replaced with pyroglutamate to improve stability during manufacture, without affecting function of the peptide. The data from further testing of these peptides (Example 5) is typically obtained using peptides where such replacement has taken place.

Table 8

| Peptide | Parent molecule | Sequence | Residues in parent | SEQ ID NO: |
|---|---|---|---|---|
| HDM01 | Der p 1 | IDLRQMRTVTPIR | 112-124 | ***1*** |
| HDM01A | Der p 1 | IDLRQMRTVTPIRMQGGSG | 112-130 | 48 |
| HDM02A | Der p 1 | RTVTPIRMQGGSG | 118-130 | 49 |
| HDM02B | Der p 1 | RTVTPIRMQGGEG | 118-130 | 50 |
| HDM03D | Der p 1 | RNQSLDLAEQELVDSASQH | 149-167 | 51 |
| HDM03E | Der p 1 | RNQSLDLAEQELVDEASQH | 149-167 | 52 |
| HDM03F | Der p 1 | RNQSLDLAEQELVDSAS | 149-165 | 53 |
| HDM03G | Der p 1 | QSLDLAEQELVDEASQHG | 151-168 | 89 |
| HDM03H | Der p 1 | LDLAEQELVDEASQHG | 153-168 | 90 |
| HDM03J | Der p 1 | LAEQELVDEASQHG | 155-168 | 91 |
| HDM03K | Der p 1 | EQELVDEASQHG | 157-168 | 92 |
| HDM03L | Der p 1 | ELVDEASQHG | 159-168 | 93 |
| HDM03M | Der p 1 | RNQSLDLAEQELVDCASQHG | 149-168 | 94 |
| HDM03N | Der p 1 | RNQSLDLAEQELVDĈASQHG | 149-168 | 95 |
| HDM03P | Der p 1 | SAYLAHRNQSLDLAEQELVDCAS | 143-166 | 96 |
| HDM03R | Der p 1 | QSLDLAEQELVDSASQHG | 151-168 | 97 |
| HDM03S | Der p 1 | LDLAEQELVDSASQHG | 153-168 | 98 |
| HDM03T | Der p 1 | LAEQELVDSASQHG | 155-168 | 99 |
| HDM03V | Der p 1 | EQELVDSASQHG | 157-168 | 100 |
| HDM03W | Der p 1 | ELVDSASQHG | 159-168 | 101 |
| HDM06A | Der p 1 | RYVAREQSSRRP | 193-205 | 54 |
| HDM06B | Der p 1 | RYVAREQSERRP | 193-205 | 55 |
| HDM07 | Der p 1 | PNVNKIREALAQT | 220-232 | 56 |
| HDM09A | Der p 1 | REALAQTHSAIAVI | 226-239 | ***5*** |
| HDM19A | Der p 2 | DQVDVKDSANHEIKK | 18-32 | 57 |
| HDM19B | Der p 2 | DQVDVKDEANHEIKK | 18-32 | 58 |
| HDM20A | Der p 2 | IIHRGKPFQLEA | 45-56 | 59 |
| HDM20B | Der p 2 | SIIHRGKPFQLEA | 44-56 | 60 |
| HDM21 | Der p 2 | KPFQLEAVFEANQNT | 50-64 | ***8*** |
| HDM21A | Der p 2 | KPFQLEAVFEANQNTK | 50-65 | ***9*** |
| HDM21B | Der p 2 | RGKPFQLEAVFEANQNT | 48-64 | ***10*** |
| HDM22A | Der p 2 | EAVFEANQNTKTAK | 55-68 | ***11*** |
| HDM23B | Der p 2 | GLEVDVPGIDPNA | 77-86 | 61 |
| HDM23C | Der p 2 | GLEVDVPGIDPNASH | 77-88 | 62 |

| | | | | |
|---|---|---|---|---|
| HDM26B | Der p 2 | GVLASAIATHAKIR | 132-145 | 63 |
| HDM26C | Der p 2 | GVLABAIATHAKIR | 132-145 | 64 |
| HDM30 | Der p 7 | DKFERHIGIIDLK | 56-68 | ***14*** |
| HDM32 | Der p 7 | IDLKGELDMRNIQ | 65-77 | ***15*** |
| HDM33 | Der p 7 | LDMRNIQVRGLKQ | 71-83 | ***16*** |
| HDM34 | Der p 7 | RNIQVRGLKQMKRVG | 74-88 | ***17*** |
| HDM35A | Der p 7 | RGLKQMKRVGDANV | 79-92 | 65 |
| HDM39A | Der p 7 | HDDVVSMEYDLAYKL | 108-121 | ***19*** |
| HDM39B | Der p 7 | HDDVVSMEYDLAYKLGDLH | 108-125 | 66 |
| HDM40A | Der p 7 | VSMEYDLAYKLGDLH | 112-124 | ***20*** |
| HDM40B | Der p 7 | VSMEYDLAYKLGDL | 112-123 | 67 |
| HDM48 | Der p 7 | TAIFQDTVRAEMTK | 187-200 | ***21*** |
| HDM48A | Der p 7 | TAIFQDTVRAEMTKVLAP | 187-204 | 68 |
| HDM49 | Der p 7 | DTVRAEMTKVLAP | 192-204 | ***22*** |
| HDM51 | Der p 7 | VDFKGELAMRNIE | 65-77 | ***23*** |
| HDM51A | Der p 7 | VDFKGELAMRNIEAR | 65-79 | ***24*** |
| HDM100 | Der p 1 | RFGISNYCQIYPPNVNK | 208-224 | 69 |
| HDM100A | Der p 1 | RFGISNYSQIYPPNVNK | 208-224 | 70 |
| HDM100B | Der p 1 | RFGISNYBQIYPPNVNK | 208-224 | 71 |
| HDM101 | Der p 1 | NYCQIYPPNVNKIREA | 213-228 | 72 |
| HDM101A | Der p 1 | NYSQIYPPNVNKIREA | 213-228 | 73 |
| HDM101B | Der p 1 | NYBQIYPPNVNKIREA | 213-228 | 74 |
| HDM102 | Der p 1 | NAQRFGISNYCQI | 205-217 | 75 |
| HDM102A | Der p 1 | NAQRFGISNYSQI | 205-217 | 76 |
| HDM102B | Der p 1 | NAQRFGISNYBQI | 205-217 | 77 |
| HDM103 | Der p 2 | KGQQYDIKYTWNVPKIAP | 99-116 | 78 |
| HDM104 | Der p 2 | WNVPKIAPKSENV | 109-121 | 79 |
| HDM201 | Der p 1 | ESVKYVQSNGGAI | 52-64 | 80 |
| HDM202 | Der p 1 | DEFKNRFLMSAEAFE | 73-87 | 81 |
| HDM202D | Der p 1 | FKNRFLMSAEA | 75-85 | 102 |
| HDM202E | Der p 1 | FKNRFLMSAE | 75-84 | 103 |
| HDM202H | Der p 1 | EFKNRFLMSAE | 74-84 | 104 |
| HDM203A | Der p 1 | DLRQMRTVTPIRMQGGCGS | 113-131 | 82 |
| HDM203B | Der p 1 | DLRQMRTVTPIRMQGGSGS | 113-131 | 83 |
| HDM204 | Der p 1 | SAYLAYRNQSLDLA | 143-156 | 84 |
| HDM205 | Der p 1 | SYYRYVAREQS | 190-199 | 85 |
| HDM206 | Der p 1 | DNGYGYFAANIDLMMIEE | 296-313 | 86 |
| HDM206A | Der p 1 | NGYGYFAANIDLMM | 297-310 | 87 |
| HDM207 | Der p 7 | DMRNIQVRGLKQMKRVGD | 72-104 | 88 |

Example 5

*Cytokine release assay and selection of peptide combinations*

Cytokine secretion profiles from PBMC's are analysed in response to the peptide stimulation using the peptides from Example 3. Supernatants from the cytokine release assay are tested for the presence of 2 cytokines, IFN-γ and IL-13, using ELISA assays. Cytokine secretion profiles from PBMC's were analysed in response to the peptide stimulation using the peptides indicated. Supernatants from the cytokine release assay were tested for the presence of 2 cytokines, IFN-γ and IL-

13, using either an ELISA assay or a multiplex bead array assay.

A typical cytokine release assay requires $40x10^6$ PBMC's per subject. In more detail, 250μl of a 200μg/ml solution of the appropriate antigen or peptide concentration is distributed into the appropriate wells of 48 well plates. Plates are the incubated in a humidified 5% $CO_2$ incubator at 37°C for a maximum of 4 hours. 250μl of a $5x10^6$ cell/ml PBMC suspension is then added to each well and the plates returned to the incubator for 5 days. Following stimulation, samples of culture supernatant are harvested for testing by ELISA or multiplex bead assay according to standard protocols.

Il-13 and IFN-gamma responses to each peptide were scored as positive T cell epitopes provided the amount of cytokine produced in the well for that peptide exceeded 100 pg/ml, i.e. 100pg per $1.25x10^6$ cells. Thus, an individual was considered to have responded to a peptide if cells from that individual yielded a response greater than 100pg/ml for either Il-13 or IFN-gamma. The percentage of responders to each peptide is shown in Figure 2.

The top five peptides by percentage of individuals with an Il-13 or IFN-gamma response greater than 100 pg/ml are HDM203B, HDM201, HDM205, HDM203A and HDM202, and (SEQ ID NOS. 83, 80, 85, 82 and 81).

HDM203A and 203B are variants of the same sequence with 203B modified such that a serine replaces a cysteine (at the third residue from the C terminus) to achieve better manufacturability and stability. Thus a preferred combination of peptides should comprise at least one of these peptides or a variant thereof.

The next most potent peptides are HDM09A, HDM03D, HDM03E, HDM101, HDM101A, HDM101B (SEQ ID NOS: 5, 51, 52, 72, 73 and 74). A preferred peptide combination may typically comprise at least one additional peptide selected from this group. Of this group HDM03D and HDM03E are sequence variants of each other with serine and aminobutyric acid (respectively) replacing cysteine (at the fifth residue from the C terminus of the native sequence of Der p 1) to achieve better manufacturability and stability. These sequences are considered equivalent.

Further variants of HDM03, namely HDM03V and HDM03W (SEQ ID NO.

100 and 101) are also considered to be suitable. These variants are fragments comprising a truncation down to the last eleven or ten (respectively) C terminal residues of HDM03D. These peptides are not included in the assay described above, but on testing are considered to be at least equivalent to HDM03D (data not shown).

HDM101, HDM101A, and HDM101B are also sequence variants of each other, with HDM101A having a serine and HDM 101B having an aminobutyric acid replacing a cysteine in HDM101 (third residue from the N terminus). All three HDM101 series peptides are considered equivalent, with HDM101A or HDM101B preferred for manufacturability and stability.

Of the remainder peptides tested the following have responses in >25% of the individuals tested: HDM01 [Der p1], HDM01A [Der p1], HDM06A [Der p2], HDM07 [Der p1], HDM19A [Der p2], HDM21A [Der p2], HDM23C [Der p2], HDM26B [Der p2], HDM35A [Der p7], HDM48 [Der p7], HDM51A [Der f 7], HDM102A [Der p1], HDM204 [Der p1] and HDM206 [Der p1] (SEQ ID NOS. 1, 48, 54, 56, 57, 9, 62, 63, 65, 21, 24, 76, 84, and 86 respectively). A preferred peptide combination may typically comprise at least one additional peptide selected from this group. When considering which additional peptides to add to the mixture, representatives from this final group should preferably be chosen from epitopes drawn from Der p2 and Der p7 since the previous groups are dominated by Der p 1. HDM26B [Der p2] and HDM 35A [Der p7] are particularly preferred. Additional studies (data not shown) demonstrate that these are the best performing peptides from Der p 2 and Der p 7 respectively.

Figure 3 shows the number of individuals who respond to a core mixture of HDM201, HDM202, HDM203B and HDM205. The incremental effect of adding HDM03D and HDM101A, and the further incremental effect of adding HDM26B and HDM35A is also shown. The benefit of adding epitopes from the second and third group of peptides is clearly shown.

Importantly, adding peptides 03D,26B,35A,101A to the core mixture converted 4 individuals from non-responders to responders. It is also apparent that removing one of the peptides 201, 202, 203B or 205 from the mixture would not reduce the number of overall responders to the proposed mixtures as most people

have three or four responses to this peptide group. This is demonstrated in Figure 4, which shows similar results for a core mixture of HDM201, HDM203B and HDM205.

5

## CLAIMS

1. A composition for use in preventing or treating allergy to house dust mites by tolerisation comprising at least four polypeptides, wherein the polypeptides are independently selected from any of the following:

(i) a polypeptide of any of HDM203B (SEQ ID 83), HDM201 (SEQ ID 80), HDM205 (SEQ ID 85), HDM203A (SEQ ID 82), HDM202 (SEQ ID 81), SEQ ID NO's 1 to 79, 84, or 86 to 104 (that is any one of SEQ ID NO's. 1 to 104);or

(ii) a variant of a polypeptide according to (i), wherein said variant is a polypeptide of length 9 to 30 amino acids that comprises a region consisting of:

- any of the sequences of (i); or
- a sequence which has at least 65% homology to any of the sequences of (i) which sequence is capable of tolerising an individual to any of the sequences of (i), or

(iii) a variant of a polypeptide according to (i), wherein said variant is a polypeptide of length 9 to 30 amino acids that comprises a region consisting of a sequence that represents either:

- a fragment of any of the sequences of (i); or
- a homologue of a fragment of any of the sequences of (i),

which sequence is capable of tolerising an individual to any of the sequences of (i) and has a length of at least 9 amino acids, and wherein said homologue has at least 65% homology to any 9 contiguous amino acids in any of the sequences of (i).

2. The composition according to claim 1, wherein the composition:

a) is capable of tolerising at least 50% or at least 60% of a panel of dust mite allergic individuals in the population; and/or

b) comprises at least four polypeptides selected from claim 1(i) or variants thereof as defined in claim 1(ii) or 1(iii), and/or

c) comprises at least one further polypeptide up to a total of thirteen unique/different polypeptides, wherein the further polypeptides:

- comprise a sequence having at least 65% sequence identity to at least 9 or more contiguous amino acids in any of SEQ ID NO: 1 to 104 above not

selected in (a); and

- are 9 to 30 amino acids in length;

and/or

d) comprises up to a maximum of thirteen polypeptides.

3. The composition according to claim 1 or 2, comprising at least one polypeptide according to claim 2 (c) which is 9 to 20 or 13 to 17 amino acids in length and/or wherein said polypeptide has at least 70% sequence identity to at least 9 or more contiguous amino acids in any of SEQ ID NO: 1 to 104.

4. A composition according to any one of the preceding claims, comprising at least one polypeptide selected from a polypeptide of SEQ ID NOS: 80, 81, 82, 83 and 85, or a variant thereof as defined in claim 1(ii) or (iii).

5. A composition according to claim 4, comprising at least two, three or four polypeptides independently selected from a polypeptide of SEQ ID NOS: 80, 81, 82, 83 and 85, or a variant thereof as defined in claim 1(ii) or (iii), with the proviso that no more than one polypeptide or variant of SEQ ID NOS: 82 and 83 is selected.

6. A composition according to any one of the preceding claims, comprising at least one additional polypeptide selected from a polypeptide of any of SEQ ID NOS: 5, 51, 52, 100, 101, 72, 73 and 74, or a variant thereof as defined in claim 1(ii) or (iii);

7. A composition according to claim 6, wherein the at least one additional peptide is selected from a polypeptide of any of SEQ ID NOS: 51, 73, 100 and 101, or a variant thereof as defined in claim 1(ii) or (iii).

8. A composition according to claim 6 or 7, comprising at least one additional polypeptide selected from a polypeptide of any of SEQ ID NOS: 1, 9, 21, 24, 48, 54, 56, 57, 62, 63, 65, 76, 84 and 86, or a variant thereof as defined in claim 1(ii) or (iii).

9. A composition according to claim 8, wherein the at least one additional peptide is selected from the polypeptides of any of SEQ ID NOS: 63 and 65, or a variant thereof as defined in claim 1(ii) or (iii).

10. A composition according to any one of the preceding claims consisting of:
a) at least one of the polypeptides of SEQ ID NOS. 83 and 82, or variants thereof as defined in claim 1(ii) or (iii); and
b) at least two of the polypeptides of SEQ ID NOS. 80, 81 and 85, or variants thereof as defined in claim 1(ii) or (iii); and optionally
c) at least one of the polypeptides of SEQ ID NOS: 51, 73, 100 and 101, or a variant thereof as defined in claim 1(ii) or (iii); and
d) at least one of the polypeptides of SEQ ID NOS: 63 and 65, or a variant thereof as defined in claim 1(ii) or (iii).

11. A composition according to claim 10, wherein the at least one polypeptide of (c) is the polypeptide of SEQ ID NO. 101.

12. The composition according to any one of the preceding claims, wherein one or more of the polypeptides have one or more modifications selected from the following:

(i) N terminal acetylation;
(ii) C terminal amidation;
(iii) one or more hydrogen on the side chain amines of Arginine and/or Lysine replaced with a methylene group;
(iv) glycosylation; and
(v) phosphorylation.

13. The composition according to any one of the preceding claims wherein at least one of the peptides has been engineered to be soluble such that it comprises:

i) N terminal to the residues of the peptide which flank a T cell

epitope: one to six contiguous amino acids corresponding to the two to six contiguous amino acids immediately N terminal to said residues in the sequence of the protein from which the peptide derives; and/or

ii) C terminal to the residues of the peptide which flank a T cell epitope: one to six contiguous amino acids corresponding to the one to six contiguous amino acids immediately C terminal to the said residues in the sequence of the protein from which the peptide derives; or

iii) both N and C terminal to the residues of the peptide which flank a T cell epitope, at least one amino acid selected from arginine, lysine, histidine, glutamate and aspartate,

wherein the polypeptide has a solubility of at least 3.5mg/ml and the T cell epitope has a solubility of less than 3.5mg/ml.

14. The composition according to any one of the preceding claims wherein at least one of the peptides has been engineered to be soluble such that additionally:

i) any cysteine residues in the native sequence of the peptide are replaced with serine or 2-aminobutyric acid; and /or

ii) any hydrophobic residues in the upto three amino acids at the N or C terminus of the native sequence of the peptide, which are not comprised in a T cell epitope, are deleted; and/or

iii) any two consecutive amino acids comprising the sequence Asp-Gly in the upto four amino acids at the N or C terminus of the native sequence of the peptide, which are not comprised in a T cell epitope, are deleted.

15. The composition according to any one of the preceding claims wherein each polypeptide has a concentration in the range of 0.03 to 200 nmol/ml, 0.3 to 200 nmol/ml or 30 to 120 nmol/ml.

16. A composition for use in preventing or treating allergy to dust mites by tolerisation comprising at least three polynucleotide sequences which when expressed cause the production of a composition as defined in any one of claims 1 to 14.

17. The composition according to claim 16, wherein each polynucleotide sequence capable of expressing a different polypeptide is present in the same or different polynucleotide vectors.

18. A vector for use in preventing or treating allergy to dust mites by tolerisation comprising at least three polynucleotide sequences which each encode a different polypeptide as defined in claim 1 and optionally one or more further polynucleotide sequences which encode different polypeptides as defined in claim 2.

19. A vector for use in preventing or treating allergy to dust mites by tolerisation comprising between four and thirteen different polynucleotide sequences, which each encode a different polypeptide as defined in claim 1 or 2, wherein at least four of the polynucleotide sequences each encode a different polypeptide selected from SEQ ID NOS. 80, 81, 82, 83 and 85, with the proviso that no more than one polypeptide of SEQ ID NOS: 82 and 83 is selected.

20. A product containing between four and thirteen polypeptides, wherein at least one, two, three or four polypeptides are independently selected from a polypeptide of SEQ ID NOS: 80, 81, 82, 83 and 85, or a variant thereof as defined in claim 1(ii) or (iii), with the proviso that no more than one polypeptide or variant of SEQ ID NOS: 82 and 83 is selected, wherein each different polypeptide is for simultaneous, separate or sequential use in preventing or treating allergy to house dust mites by tolerisation.

21. A product containing between four and thirteen polynucleotide sequences, which each encode a different polypeptide as defined in claim 1 or 2, wherein at least

four of the polynucleotide sequences each encode a different polypeptide selected from SEQ ID NOS. 80, 81, 82, 83 and 85, with the proviso that no more than one polypeptide of SEQ ID NOS: 82 and 83 is selected, and wherein each different polypeptide is for simultaneous, separate of sequential use in the prevention or treatment of allergy to dust mites in a human.

22. A pharmaceutical formulation for use in preventing or treating allergy to dust mites by tolerisation comprising a composition according to any one of claims 1 to 17; a vector according to any one of claims 18 or 19; or a product according to any one of claims 20 or 21; and a pharmaceutically acceptable carrier or diluent.

23. The composition, vector or product according to claim 22, formulated for oral administration, nasal administration, epicutaneous administration, subcutaneous administration, sublingual administration, intradermal administration, buccal administration or for administration by inhalation or by injection.

24. The composition as defined in any one of claims 1 to 17 or product as defined in claim 20 or 21, additionally comprising a further polypeptide allergen for use in tolerising an individual to the further polypeptide allergen.

25. An *in vitro* method of determining whether T cells recognize a polypeptide as defined in claim 1 comprising contacting said T cells with said polypeptide and detecting whether said T cells are stimulated by said polypeptide.

26. An in vitro method of determining whether an individual has or is at risk of a condition wherein the condition is characterized by allergic symptoms in response to a house dust mite allergen, the method comprising testing whether the individual has T cells which respond to a composition as defined in any one of claims 1 to 17, thereby determining whether the individual has or is at risk of the condition.

27. A method according to claim 26 wherein a T-cell immune response to said composition is measured by contacting the composition with T cells in a sample

taken from the subject, under conditions which allow the composition and the T cells to interact; and determining whether or not any of the T cells are stimulated and thereby determining whether or not a T-cell immune response is present or absent.