



Office de la Propriété
Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2091449 C 2002/11/05

(11)(21) **2 091 449**

(12) **BREVET CANADIEN
CANADIAN PATENT**

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 1991/09/10
(87) Date publication PCT/PCT Publication Date: 1992/03/19
(45) Date de délivrance/Issue Date: 2002/11/05
(85) Entrée phase nationale/National Entry: 1993/03/10
(86) N° demande PCT/PCT Application No.: AU 1991/000417
(87) N° publication PCT/PCT Publication No.: 1992/004445
(30) Priorité/Priority: 1990/09/11 (580,655) US

(51) Cl.Int.⁵/Int.Cl.⁵ C12N 15/12, A61K 39/35, G01N 33/566,
C07K 14/435
(72) Inventeurs/Inventors:
Thomas, Wayne R., AU;
Chua, Kaw-Yan, AU
(73) Propriétaire/Owner:
Institute for Child Health Research, AU
(74) Agent: RIDOUT & MAYBEE LLP

(54) Titre : CLONAGE ET SEQUENCAGE D'ALLERGENES DE DERMATOPHAGOIDES (ACARIENS DETRITICOLES)
(54) Title: CLONING AND SEQUENCING OF ALLERGENS OF DERMATOPHAGOIDES (HOUSE DUST MITE)

(57) **Abrégé/Abstract:**

Isolated DNA encoding a protein allergen or an epitope of a protein allergen of Dermatophagoides farinae or Dermatophagoides pteronyssinus. An isolated protein allergen or peptide which includes an epitope of a protein allergen of Dermatophagoides farinae or Dermatophagoides pteronyssinus. Diagnostic and therapeutic compositions and methods of use thereof.



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

2091449



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁵:

C12N 15/12, C12P 21/02

C07K 15/08, G01N 33/68

A61K 39/35

A1

(11) International Publication Number:

WO 92/04445

(43) International Publication Date:

19 March 1992 (19.03.92)

(21) International Application Number: PCT/AU91/00417

(22) International Filing Date: 10 September 1991 (10.09.91)

(30) Priority data:

580,655

11 September 1990 (11.09.90) US

(71) Applicant: THE WESTERN AUSTRALIAN RESEARCH
INSTITUTE FOR CHILD HEALTH LTD. [AU/AU];
Princess Margaret Hospital for Children, Roberts Road,
Subiaco, W.A. 6008 (AU).(72) Inventors: THOMAS, Wayne, Robert ; 31 Taylor Road,
Nedlands, W.A. 6009 (AU). CHUA, Kaw-Yan ; 35 Mun-
ja Way, Nollamara, W.A. 6061 (AU).(74) Agents: SLATTERY, John, Michael et al.; Davies & Col-
lison, 1 Little Collins Street, Melbourne, VIC 3000 (AU).(81) Designated States: AT (European patent), AU, BE (Euro-
pean patent), CA, CH (European patent), DE (Euro-
pean patent), DK (European patent), ES (European pa-
tent), FR (European patent), GB (European patent), GR
(European patent), IT (European patent), JP, KR, LU
(European patent), NL (European patent), SE (Euro-
pean patent).

Published

With international search report.

(54) Title: CLONING AND SEQUENCING OF ALLERGENS OF DERMATOPHAGOIDES (HOUSE DUST MITE)

(57) Abstract

Isolated DNA encoding a protein allergen or an epitope of a protein allergen of *Dermatophagoides farinae* or *Dermatophagoides pteronyssinus*. An isolated protein allergen or peptide which includes an epitope of a protein allergen of *Dermatophagoides farinae* or *Dermatophagoides pteronyssinus*. Diagnostic and therapeutic compositions and methods of use thereof.

```
GAATTCGGTTTCTTCATCAAAATTAAATTCATCAAAA ATG AAA TTC GTT TTG GCC ATT 62
Met Lys Phe Val Leu Ala Ile
-90
GCC TCT TTG TTG GTA TTG ACC ACT GTT TAT GCT CGT CCA GCT TCA ATC AAA ACT 116
Ala Ser Leu Leu Val Leu Ser Thr Val Thr Ala Arg Pro Ala Ser Ile Lys Thr
-70
TTT GAA GAA TTC AAA AAA GCC TTC AAC AAA AAC TAT GCC ACC GTT GAA GAG GAA 170
Phe Glu Glu Phe Lys Lys Ala Phe Asn Lys Asn Tyr Ala Thr Val Glu Glu Glu
-50
GAA GTT CCC GGT AAA AAC TTT TTG GAA TCA TTG AAA TAT GTT GAA GCT AAC AAA 224
Glu Val Ala Arg Lys Asn Phe Leu Glu Ser Leu Lys Tyr Val Glu Ala Asn Lys
-30
GCT GCC ATC AAC CAT TTG TCC GAT TTG TCA TTG GAT GAA TTC AAA AAC GGT TAT 278
Gly Ala Ile Asn Ala Leu Ser Asp Leu Ser Leu Asp Glu Phe Lys Asn Arg Tyr
-10
TTG ATG AGT GCT GAA GCT TTT GAA CAA CTC AAA ACT CAA TTC GAT TTG AAT GCC 332
Leu Met Ser Ala Glu Ala Phe Glu Glu Leu Lys Thr Glu Phe Asp Leu Asn Ala
-1
GAA ACA AGC GCT TCC GGT ATC AAT TCG GTT AAC GTT CCA TCG GAA TTG GAT TTA 386
Glu Thr Ser Ala Cys Arg Ile Asn Ser Val Asn Val Pro Ser Glu Leu Asp Leu
20
CCA TCA CTG CGA ACT CTC ACT CCA ATC GGT ATG CAA CGA GGC TGT GGT TCA TGT 440
Arg Ser Leu Arg Thr Val Thr Pro Ile Arg Met Glu Gly Gly Cys Gly Ser Cys
40
TGG GCT TTC TCT GGT GTT GCC GCA ACT GAA TCA GCT TAT TTG GCT TAC GGT AAC 494
Trp Ala Phe Ser Gly Val Ala Ala Thr Glu Ser Ala Tyr Leu Ala Tyr Arg Asn
60
ACG TCT TTG GAT CTT TCT GAA CAG GAA CTC GTC GAT TCC GCA TCT CAA CAC GCA 548
Thr Met Leu Asp Leu Ser Glu Glu Glu Leu Val Asp Cys Ala Ser Glu Ala Gly
80
TGT CAC GGC GAT ACA ACA CCA ACA GGC ATC GAA TAC ATC CAA CAA AAT GGT GTC 602
Cys His Gly Asp Thr Ile Pro Arg Gly Ile Glu Tyr Ile Glu Glu Asn Gly Val
100
GTT GAA GAA ACA AGC TAT CCA TAC GGT GCA GCA GAA CAA CAA TCC CCA GCA CCA 656
Val Glu Glu Arg Ser Tyr Pro Tyr Val Ala Arg Glu Glu Arg Cys Arg Arg Pro
120
AAT TCG CAA CAT TAC GGT ATC TCA AAC TAC TCC CAA ATT TAT CCA CCA GAT GTG 710
Asn Ser Glu His Tyr Gly Ile Ser Asn Tyr Cys Glu Ile Tyr Pro Pro Asp Val
140
AAA CAA ATC GGT GAA GCT TTG ACT CAA ACA CAC ACA GGT ATT GCT GTC ATT ATT 764
Lys Glu Ile Arg Glu Ala Leu Thr Glu Thr His Thr Ala Ile Ala Val Ile Ile
160
GGC ATC AAA GAT TTT ACA GCT TTC CAA CAT TAT GAT GCA GCA ACA ATC ATT CAA 818
Gly Ile Lys Asp Leu Arg Ala Phe Glu His Tyr Asp Gly Arg Thr Ile Ile Glu
180
CAT GAC AAT GGT TTT CAA CCA AAC TAT CAT GCT GTC AAC ATT GTC GGT TAC GCA 872
His Asp Asn Gly Tyr Glu Pro Asn Tyr His Ala Val Asn Ile Val Gly Tyr Gly
200
AGT ACA CAA GGC GAC GAT TAT TCG ATC GTA CCA AAC ACT TCG GAT ACT ACC TCG 926
Ser Thr Glu Gly Asp Asp Tyr Trp Ile Val Arg Asn Ser Trp Asp Thr Thr Trp
220
GCA GAT AGC GCA TAC GCA TAT TTC CAA GCC GCA AAC AAC CTC ATG ATC ATC CAA 980
Gly Asp Ser Gly Tyr Gly Tyr Phe Glu Ala Gly Asn Asn Leu Met Met Ile Glu
240
CAA TAT CCA TAT GTT GTA ATC ATG TGAACATTCGAATTCGAATATATTTATTTTTCATCAAA 1044
Gln Tyr Pro Tyr Val Val Ile Met
AATACAACTACTCTTCGAGTATTTTACTCGGAATTC 1084
```

2091449

5

CLONING AND SEQUENCING OF ALLERGENS
OF DERMATOPHAGOIDES (HOUSE DUST MITE)

10

Description

Background

15

Recent reports have documented the importance of responses to the Group I and Group II allergens in house dust mite allergy. For example, it has been documented that over 60% of patients have at least 50% of their anti-mite antibodies directed towards these proteins (Lind, P. et.al., Allergy, 39:259-274 (1984); van der Zee, J.S. et.al., J.Allergy Clin. Immunol., 81:884-896 (1988)). It is possible that children show a greater degree of reactivity (Thompson, P.J. et.al., Immunology,

20

2091449

64:311-314 (1988)). Allergy to mites of the genus Dermatophagoides (D.) is associated with conditions such as asthma, rhinitis and ectopic dermatitis. Two species, D. pteronyssinus and D. farinae, predominate and, as a result, considerable effort has been expended in trying to identify the allergens produced by these two species. D. pteronyssinus mites are the most common Dermatophagoides species in house dust in Western Europe and Australia. The species D. farinae predominates in other countries, such as, North America and Japan (Wharton, G.W., J. Medical Entom., 12:577-621 (1976)). It has long been recognized that allergy to mites of this genus is associated with diseases such as asthma, rhinitis and atopic dermatitis. It is still not clear what allergens produced by these mites are responsible for the allergic response and associated conditions.

Summary of the Invention

20 The present invention relates to isolated DNA which encodes a protein allergen of Dermatophagoides (D.) house dust mite) or a peptide which includes at least one epitope of a protein allergen of a house dust mite of the genus Dermatophagoides. It particularly relates to DNA encoding major allergens of the species D. farinae, designated Der_f I and Der_f II, or portions of these major allergens (i.e., peptides which include at least one epitope of Der_f I or of Der_f II). It also particularly relates to DNA encoding major allergens of D. pteronyssinus, designated Der_p I and Der_p II, or

2091449

portions of these major allergens (i.e., peptides which include at least one epitope of Der_p I or of Der_p II.

The present invention further relates to
5 proteins and peptides encoded by the isolated Dermatophagoides (e.g., D. farinae, D. pteronyssinus) DNA. Peptides of the present invention include at least one epitope of a D. farinae allergen (e.g., at least one epitope of Der_f I or of Der_f II) or at least one epitope of a D. pteronyssinus
10 allergen (e.g., at least one epitope of Der_p I or of Der_p II). It also relates to antibodies specific for D. farinae proteins or peptides and to antibodies specific for D. pteronyssinus proteins or
15 peptides.

Dermatophagoides DNA, proteins and peptides of the present invention are useful for diagnostic and therapeutic purposes. For example, isolated D. farinae protein or peptide can be used to detect
20 sensitivity in an individual to house dust mites and can be used to treat sensitivity (reduce sensitivity or desensitize) in an individual, to whom therapeutically effective quantities of the D. farinae protein or peptide is administered. For example,
25 isolated D. farinae protein allergen, such as Der_f I or Der_f II can be administered periodically, using standard techniques, to an individual in order to desensitize the individual. Alternatively, a peptide which includes at least one epitope of Der_f I or of Der_f II can be administered for this
30 purpose. Isolated D. pteronyssinus protein allergen, such as Der_p I or Der_p II, can be administered as described for Der_f I or Der_f II.

2091449

Similarly a peptide which includes at least one Der
p I epitope or at least one Der_p II epitope can be
administered for this purpose. A combination of
these proteins or peptides (e.g., Der_f I and Der_f
5 II; Der_p I and Der_p II; or a mixture of both Der_f
and Der_p proteins) can also be administered. The
use of such isolated proteins or peptides provides a
means of desensitizing individuals to important
house dust mite allergens.

10 Brief Description of the Drawings

Figure 1 is a restriction map of the cDNA
insert of clone λ gt11 f 1, including a schematic
representation of the strategy of DNA sequencing.
Arrows indicate directions in which sequences were
15 read.

Figure 2 is the nucleotide sequence and the
predicted amino acid sequence of cDNA λ gt11 f 1.
Numbers above are nucleotide positions; numbers to
the left are amino acid positions. Positive amino
20 acid residue numbers correspond to the sequence of
the mature excreted Der_f I beginning with
threonine. Negative sequence numbers refer to the
signal peptide and the proenzyme regions of Der_f I.
The arrows indicate the beginning of the proenzyme
25 sequence and the mature
Der_f I, respectively. The underlined residues -81
to -78 make up the proposed cleavage site for the
proenzyme formation, while the underlined residues
53-55 represent a potential N-glycosylation site.
30 The termination TGA codon and the adjacent polya-
denylation signal are also underlined. Amino acid

-5-

residues 1-28 correspond to a known tryptic peptide sequence determined by conventional amino acid sequencing analysis.

Figure 3 is a composite alignment of the amino acid sequences of the mature Der_p I and Der_f I proteins. The numbering above the sequence refers to Der_p I. The asterisk denotes the gap that was introduced for maximal alignment. The symbol (.) is used to indicate that the amino acid residue of Der_f I at that position is identical to the corresponding amino acid residue of Der_p I. The arrows indicate those residues making up the active site of Der_p I and Der_f I.

Figure 4 is a comparison of the amino acid sequence in the pre- and pro-peptide regions of Der_f I with those of rat cathepsin H, rat cathepsin L, papain, aleurain, CP1, CP2, rat cathepsin B, CTLA-2, MCP, Der_p I and actinidin. Gaps, denoted by dashes, were added for maximal alignment. Double asterisks denote conserved amino acid residues which are shared by greater than 80% of the proenzymes; single asterisks show residues which are conserved in greater than 55% of the sequences. The symbol (.) is used to denote semiconserved equivalent amino acids which are shared by greater than 90% of the proenzyme regions.

Figure 5 is a hydrophilicity plot of the Der_p I mature protein and a hydrophilicity plot of the Der_f I mature protein produced using the Hopp-Woods algorithm computed with the MacTM Vector Sequence Analysis Software (IBI, New Haven) using a 6 residue window. Positive values indicate relative

2091449

hydrophilicity and negative values indicating relative hydrophobicity.

Figure 6 is the nucleotide sequence and the predicted amino acid sequence of Der_f II cDNA. Numbers to the right are nucleotide positions and numbers above are amino acid residues. The stop (TAA) signal is underlined. The first 8 nucleotides are from the oligonucleotide primer used to generate the cDNA, based on the Der_p II sequence.

Figure 7 is the nucleotide sequence and predicted amino acid sequence of cDNA λ gt11 pl(13T). Numbers to the right are nucleotides and numbers above are amino acid positions. Positive amino acids correspond to the sequence of mature excreted Der_p I beginning with threonine.

Figure 8 is a restriction map of Der_f II cDNA, which was generated by computer from the sequence data. A map of Der_p II similarly generated is shown for comparison. There are few common restriction enzyme sites conserved. Sites marked with an asterisk were introduced by cloning procedures.

Figure 9 shows the alignment of Der_f II and Der_p II cDNA sequences. Numbers to the right are nucleotide position and numbers above and amino acid residues. The top line gives the Der_p II nucleotide sequence and the second the Der_p II amino acid residues. The next two lines show differences of Der_f II to these sequences.

Figure 10 is a hydrophilicity plot of Der_f II and Der_p II using the Hopp-Woods algorithm computed with the Mac Vector Sequence Analysis Software (IBI, New Haven) using a 6-residue window.

2091449

Figure 11 is a composite alignment of the amino acid sequences of five Der p I clones (a)-(e) which illustrates polymorphism in the Der p I protein. The numbering refers to the sequence of the Der p I(a) clone. The symbol (-) is used to indicate that the amino acid residue of a Der p I clone is identical to the corresponding amino acid residue of Der p I(a) at that position. The amino acid sequences of these clones indicates that there may be significant variation in Der p I, with five polymorphic amino acid residues found in the five sequences.

2091449

Detailed Description of the Invention

The present invention relates to a nucleotide sequence coding for an allergen from the house dust mite Dermatophagoides and to the encoded

5 Dermatophagoides protein or peptide which includes at least one epitope of the Dermatophagoides allergen. It particularly relates to a nucleotide sequence capable of expression in an appropriate host of a major allergen of D. farinae, such as Der

10 f I or Der_f II, or of a peptide which includes at least one epitope of Der_f I or of Der_f II. It also particularly relates to a nucleotide sequence capable of expression in an appropriate host of a major allergen of D. pteronyssinus, such as Der_p I

15 or Der_p II, or of a peptide which includes at least one epitope of Der_p I or of Der_p II. The Dermatophagoides nucleotide sequence is useful as a probe for identifying additional nucleotide sequences which hybridize to it and encode other

20 mite allergens, particularly D. farinae or D. pteronyssinus allergens. Further, the present invention relates to nucleotide sequences which hybridize to a D. farinae protein-encoding nucleotide sequence or a D. pteronyssinus

25 protein-encoding nucleotide sequence but which encode a protein from another species or type of house dust mite, such as D. microceras (e.g., Der_m I and Der_m II).

The encoded Dermatophagoides mite allergen or

30 peptide which includes at least one Dermatophagoides (Der_f I or Der_f II; Der_p I or Der_p II) epitope can be used for diagnostic purposes (e.g., as an

2091449

antigen) and for therapeutic purposes (e.g., to desensitize an individual). Alternatively, the encoded house dust mite allergen can be a protein or peptide, such as a D. microceras protein or peptide, which displays the antigenicity of or is cross-reactive with a Der_f or a Der_p allergen; generally, these have a high degree of amino acid homology.

Accordingly, the present invention also relates to compositions which include a Dermatophagoides allergen (e.g., Der_f I or allergen, Der_f II allergen; Der_p I or Der_p II allergen or other D. allergen cross-reactive therewith) or a peptide which includes at least one epitope of a Dermatophagoides allergen (Der_f I, Der_f II, Der_p I, Der_p II or other D. allergen cross-reactive therewith) individually or in combination, and which can be used for therapeutic applications (e.g., desensitization). As is described below, DNA coding for major allergens from house dust mites have been isolated and sequenced. In particular, and as is described in greater detail in the Examples, two cDNA clones, coding, respectively, for Der_f I allergen and Der_f II allergen, have been isolated and sequenced. The nucleotide sequences of Der_p I and Der_p II have also been determined. The nucleotide sequence of each of these clones has been compared with that of the homologous allergen from the related mite D. pteronyssinus (Der_p I and Der_p II, respectively), as has the predicted amino acid sequence of each.

The following is a description of isolation and sequencing of the two cDNA clones coding for Der_f

2091449⁻¹⁰

allergens and their comparison with the
corresponding D. pteronyssinus allergen and a
description of use of the nucleotide sequences and
encoded products in a diagnostic or a therapeutic
5 context.

Isolation and Sequence Analysis of Der_f I

A cDNA clone coding for Der_f I, a major
allergen from the house dust mite D. farinae, has
been isolated and sequenced. A restriction map of
10 the cDNA insert of the clone is represented in
Figure 1, as is the strategy of DNA sequencing.
This Der_f I cDNA clone contains a 1.1-kb cDNA
insert encoding a typical signal peptide, a pro-
enzyme region and the mature Der_f I protein. The
15 product is 321 amino acid residues: a putative 18
residue signal peptide, an 80 residue proenzyme
(pro-peptide) region, and a 223 residue mature
enzyme region. The derived molecular weight is
25,191. The nucleotide sequence and the predicted
20 amino acid sequence of the Der_f I cDNA are repre-
sented in Figure 2. The deduced amino acid sequence
shows significant homology to other cysteine pro-
teases in the pro-region, as well as in the mature
protein. Sequence alignment of the mature Der_f I
25 protein with the homologous allergen Der_p I from
the related mite D. pteronyssinus (Figure 3)
revealed a high degree of homology (81%) between the
two proteins, as predicted by previous sequencing at
the protein level. In particular, the residues com-
30 prising the active site of these enzymes were
conserved and a potential N-glycosylation site was

present at equivalent positions in both mite allergens.

Conserved cysteine residue pairs (31, 71) and (65, 103), where the numbering refers to Der_p I, are apparently involved in disulphide bond formation on the basis of the assumed similarity of the three dimensional structure of Der_p I and Der_f I to that of papain and actinidin, which also have an additional disulphide bridge. The fifth and final cysteine residue for which there is a homologous cysteine residue in papain and actinidin is the active site cysteine (residue 35 in Der_f I). It is not unlikely that the two extra cysteine residues present in Der_p I and Der_f I may be involved in forming a third disulphide bridge.

The potential N-glycosylation site in Der_p I is also present at the equivalent position in Der_f I, with conservation of the crucial first and last residues of the tripeptide site. The degree of glycosylation of Der_f I and Der_p I has yet to be determined. Carbohydrates, including mannose, galactose, N-acetylglucosamine and N-acetylgalactosamine, have been reported in purified preparations of these mite allergens (Chapman, M.D., J. Immunol., 125:587-592 (1980); Wolden, S. et al., Int. Arch. Allergy Appl. Immunol., 68:144-151 (1982)).

Given the degree of homology over the first thirty N-terminal amino acid residues between mature Der_p I and Der_m I (70%) and mature Der_f I and Der_m I (97%) with the Der_m I residues determined by conventional amino acid sequencing (Platts-Mills TAE et al., In: Mite Allergy, a World-Wide Problem,

2091449

27-29 (1988); Lind, P. and N. Horn, In: Mite Allergy, a World-Wide Problem, 30-34 (1988)), it is probable that the full mature Der_m I sequence will confirm an overall 70-80% homology between the Group I mite allergens. Der_m I is an allergen from D. microceras. High homology between the proenzyme moieties of Der_p I and Der_f I (91%) over the residues -23 to -1 and the structural analysis of Der_f I suggests that the Group I allergens are likely to have N-terminal extension peptides of the mature protein of homologous structure and, at least for the pro-peptide, composition.

Studies on the fine structure of the design of signal sequences have identified three structurally dissimilar regions so far: a positively charged N-terminal (n) region, a central hydrophobic (h) region and a more polar C-terminal (c) region that seems to define the cleavage site (Von Heijne, G., EMBO J., 3:2315-2323 (1984); Eur. J. Biochem., 133:17-21 (1983); J. Mol. Biol., 184:99-105 (1985)). Analysis of the signal peptide of Der_f I revealed that it, too, contained these regions (Figure 4). The n-region is extremely variable in length and composition, but its net charge does not vary appreciably with the overall length, and has a mean value of about +1.7. The n-region of the Der_f I signal peptide, with a length of two residues, has a net charge of +2 contributed by the initiator methionine (which is unformylated and hence positively charged in eukaryotes) and the adjacent lysine (Lys) residue. The h-region of Der_f I is enriched with hydrophobic residues, the

2091449

characteristic feature of this region, with only one hydrophilic residue serine (Ser) present which can be tolerated. The overall amino acid composition of the Der_f I c-region is more polar than that of the h-region as is found in signal sequences with the h/c boundary located between residues -6 and -5, which is its mean position in eukaryotes. Thus, the Der_f I pre-peptide sequence appears to fulfill the requirements to which a functional signal sequence must conform.

While the signal sequence of Der_f I and other cysteine proteases share structural homology, all being composed of the n,h and c-regions, they are highly variable with respect to overall length and amino acid sequence, as is clear in Figure 4. However, significant sequence homology has been shown between the pro-regions of cysteine protease precursors (Ishidoh, K. et al., FEBS Letters, 226:33-37 (1987)). Alignment of the proenzyme regions of Der_f I and a number of other cysteine proteases (Figure 4) indicated that these proregions share a number of very conserved residues as well as semi-conserved residues which were present in over half of the sequences. This homology was increased if conservative amino acids such as valine (Val), isoleucine (Ile) and leucine (Leu) (small hydrophobic residues) or arginine (Arg) and Lys (positively charged residues) were regarded as identical. The Der_f I proregion possessed six out of the seven highly conserved amino acids and all the residues at sites of conservative changes. The homology at less conserved sites was lower. Homology in the pro-

2091449

peptide, in particular the highly conserved residues, may be important when considering the function of the pro-peptide in the processing of these
5 enzymes, since it indicates that these sequences probably have structural and functional similarities.

Highly cross-reactive B cell epitopes on Der_f I and Der_p I have been demonstrated using antibodies present in mouse, rabbit and human sera
10 (Heymann, P.W. et al., J. Immunol. 137:2841-2847 (1986); Platts-Mills, TAE et al., J. Allergy Clin. Immunol., 78:398-407 (1986)). However, species-specific epitopes have also been defined in these
15 systems. Murine monoclonal antibodies bound predominantly to species-specific determinants (Platts-Mills TAE et al., J. Immunol., 139:1479-1484 (1987)). Some 40% of rabbit anti-Der_p I reactivity was accounted for by epitopes unique to Der_p I
20 (Platts-Mills, TAE et al., J. Allergy Clin. Immunol., 78:398-407 (1986)), and some species-specific binding of antibodies from allergic humans was observed, although the majority bind to cross-reactive epitopes (Platts-Mills TAE et al., J.
25 Immunol., 139:1479-1484 (1987)).

The recombinant DNA strategy of gene fragmentation and expression was used (Greene, W.K. et al., Immunol. (1990)) to define five antigenic regions of recombinant Der_p I which contained B cell epitopes
30 recognized by a rabbit anti-Der_p I antiserum. Using the technique of immunoabsorption, three of these putative epitopes were shown to be shared with

2091449

Der_f I (located on regions containing amino acid residues 34-47, 60-72 and 166-194) while two appeared to be specific for Der_p I (regions 82-99 and 112-140). Differences in the reactivity of these peptides to rabbit anti-D.farinae supported the above division into cross-reactive and species-specific epitopes. The sequence differences shown between the Der_p I and the Der_f I proteins are primarily located in the N and C terminal regions, as well as in an extended surface loop (residues 85-136) linking the two domains of the enzyme that includes helix D (residues 127-136), as predicted from the secondary and tertiary structures of papain and actinidin (Baker, E.N. and J. Drenth, In: Biological Macromolecules and Assemblies, Vol. 3, pp. 314-368, John Wiley and Sons, NY. (1987)). The surface location of these residues is supported by the hydrophilicity plots of Der_p I and Der_f I in Figure 5, which illustrate the predominantly hydrophilic nature of this region that predicts surface exposure. This region also contains the two species-specific B cell epitopes recognized by the rabbit anti-Der_p I serum (see above). Analysis of the sequences in the regions containing the cross-reactive epitopes showed that two of the cross-reactive epitopes (located in regions 34-47 and 60-72) are completely conserved between Der_p I and Der_f I, while the majority of residues in a third cross-reactive epitope-containing region (residues region 166-194) were conserved.

2091449

Expression of results in production of pre-pro-Der_f I protein in E. coli a recombinant protein of greater solubility, stability and antigenicity than that of recombinant Der_p I. Protein encoded
5 by Der_f I cDNA has been expressed using a pGEX vector and has been shown by radioimmune assay to react with rabbit anti-D. farinae antibodies. The availability of high yields of soluble Der_f I allergen and antigenic derivatives will facilitate
10 the development of diagnostic and therapeutic agents and the mapping of B and T cell antigenic determinants.

With the availability of the complete amino acid sequence of recombinant Der_f I, mapping of the
15 epitopes recognized by both the B and T cell compartments of the immune system can be carried out. The use of techniques such as the screening of overlapping synthetic peptides, the use of monoclonal antibodies and gene fragmentation and
20 expression should enable the identification of both the continuous and topographical epitopes of Der_f I. It will be particularly useful to determine whether allergenic (IgE-binding) determinants have common features and are intrinsically different from
25 antigenic (IgG-binding) determinants and whether T cells recognize unique epitopes different from those recognized by B cells. Studies to identify the Der_f I epitopes reactive with mite allergic human IgE antibodies and the division of these into
30 determinants cross-reactive with Der_p I and determinants unique to Der_f I can also be carried out. B cell (and T cell) epitopes specific for

either species can be used to provide useful diagnostic reagents for determining reactivity to the different mite species, while cross-reacting epitopes are candidates for a common immunotherapeutic agent.

As described in U.S. patent no. 5,972,352, the molecular cloning of mite allergens resulted in the isolation of a cDNA clone coding for Der_p I which contained a 0.8 - kb CdNA insert. Sequence analysis revealed that the 222 amino acid residue mature recombinant Der_p I protein showed significant hemology with a group of cysteine proteases, including actinidin, papain, cathepsin H and cathepsin B.

Isolation and Sequence Analysis of Der_f II

A cDNA clone coding for Der_f II, a major allergen from the house dust mite D. farinae, has been isolated and sequenced, as described in Example 2. The nucleotide sequence and the predicted amino acid sequence of the Der_f II cDNA are represented in Figure 6. A restriction map of the cDNA insert of a clone coding for Der_f II is represented in Figure 8.

Figure 9 shows the alignment of Der_f II and Der_p II cDNA sequences.

The homology of the sequence of Der_f II with Der_p II (88%) is higher than the 81% homology found with Der_p I and Der_f I, which is significantly different ($p < 0.05$) using the χ^2 distribution. The reason for this may simply be that the Group I

2091449

allergens are larger and each residue may be less critical for the structure and function of the molecule. It is known, for example, that assuming they adopt a similar conformation to other cysteine proteases, many of the amino acid differences in Der
5 p I and Der_f I lie in residues linking the two domain structures of the molecules. The 6 cysteine molecules are conserved between the group II allergens, suggesting a similar disulphide bonding,
10 although this may be expected, given the high overall homology. Another indication of the conservation of these proteins is that 34/55 of the nucleotide changes of the coding sequence are in the third base of a codon, which usually does not change
15 the amino acid. Residues that may be of importance in the function of the molecule are Ser 57 where all three bases are changed but the amino acid is conserved. A similar phenomenon exists at residue 88, where a complete codon change has conserved a
20 small aliphatic residue. Again, like Der_p II, the Der_f II cDNA clone does not have a poly A tail, although the 3' non-coding region is rich in adenosine and has two possible polyadenylation signals ATAA. The nucleotides encoding the first
25 four residues are from the PCR primer which was designed from the known homology of Der_p II and Der_f II from N-terminal amino acid sequencing. A primer based on the C-terminal sequence can now be used to determine these bases, as well as the signal
30 sequence.

2091449

Uses of the subject allergenic proteins/peptides and
DNA encoding same

The materials resulting from the work described herein, as well as compositions containing these materials, can be used in methods of diagnosing, treating and preventing allergic responses to mite allergens, particularly to mites of the genus Dermatophagoides, such as D. farinae and other species (e.g., D. pteronyssinus and D. microceras). In addition, the cDNA (or the mRNA from which it was transcribed) can be used to identify other similar sequences. This can be carried out, for example, under conditions of low stringency and those sequences having sufficient homology (generally greater than 40%) can be selected for further assessment using the method described herein. Alternatively, high stringency conditions can be used. In this manner, DNA of the present invention can be used to identify sequences coding for mite allergens having amino acid sequences similar to that of Der_f I, Der_f II, Der_p I or Der_p II. Thus, the present invention includes not only Der_f and Der_p II, but other mite allergens as well (e.g., other mite allergens encoded by DNA which hybridizes to DNA of the present invention).

Proteins or peptides encoded by the cDNA of the present invention can be used, for example, as "purified" allergens. Such purified allergens are useful in the standardization of allergen extracts or preparations which can be used as reagents for the diagnosis and treatment of allergy to house dust mites. Through use of the peptides of the present

2091449

invention, allergen preparations of consistent, well-defined composition and biological activity can be made and administered for therapeutic purposes (e.g., to modify the allergic response of a house
5 dust mite-sensitive individual). Der_f I or Der_f II peptides or proteins (or modified versions thereof, such as are described below) may, for example, modify B-cell response to Der_f I or Der_f II, T-cell response to Der_f I and Der_f II, or both
10 responses. Similarly, Der_p I or Der_p II proteins or peptides may be used to modify B-cell and/or T-cell response to Der_p I or Der_p II. Purified allergens can also be used to study the mechanism of immunotherapy of allergy to house dust mites,
15 particularly to Der_f I, Der_f II, Der_p I and Der_p II, and to design modified derivatives or analogues which are more useful in immunotherapy than are the unmodified ("naturally-occurring") peptides.

In those instances in which there are epitopes
20 which are cross-reactive, such as the three epitopes described herein which are shared by Der_f I and Der_p I, the area(s) of the molecule which contain the cross-reactive epitopes can be used as common immunotherapeutic peptides to be administered in
25 treating allergy to the two (or more) mite species which share the epitope. For example, the cross-reactive epitopes could be used to induce IgG blocking antibody against both allergens (e.g., Der_f I and Der_p I allergen). A peptide containing a
30 univalent antibody epitope can be used, rather than the entire molecule, and may prove advantageous because the univalent antibody epitope cannot

2091449

crosslink mast cells and cause adverse reactions during desensitizing treatments. It is also possible to attach a B cell epitope to a carrier molecule to direct T cell control of allergic responses.

Alternatively, it may be desirable or necessary to have peptides which are specific to a selected Dermatophagoides allergen. As described herein, two epitopes which are apparently Der_p I-specific have been identified. A similar approach can be used to identify other species-specific epitopes (e.g., Der_p I or II, Der_f I or II. The presence in an individual of antibodies to the species-specific epitopes can be used as a quick serological test to determine which mite species is causing the allergic response. This would make it possible to specifically target therapy provided to an individual to the causative species and, thus, enhance the therapeutic effect.

Work by others has shown that high doses of allergens generally produce the best results (i.e., best symptom relief). However, many people are unable to tolerate large doses of allergens because of allergic reactions to the allergens. Modification of naturally-occurring allergens can be designed in such a manner that modified peptides or modified allergens which have the same or enhanced therapeutic properties as the corresponding naturally-occurring allergen but have reduced side effects (especially, anaphylactic reactions) can be produced. These can be, for example, a peptide of the present invention (e.g., one having all or a

2091449

portion of the amino acid sequence of Der_f I or Der_f II, Der_p I or Der_p II). Alternatively, a combination of peptides can be administered. A modified peptide or peptide analogue (e.g., a

5 peptide in which the amino acid sequence has been altered to modify immunogenicity and/or reduce allergenicity or to which a component has been added for the same purpose) can be used for desensitization therapy.

10 Administration of the peptides of the present invention to an individual to be desensitized can be carried out using known techniques. A peptide or combination of different peptides can be administered to an individual in a composition which

15 includes, for example, an appropriate buffer, a carrier and/or an adjuvant. Such compositions will generally be administered by injection, inhalation, transdermal application or rectal administration. Using the information now available, it is possible

20 to design a Der_f I or Der_f II peptide which, when administered to a sensitive individual in sufficient quantities, will modify the individual's allergic response to a Der_f I and/or Der_f II. This can be done, for example, by examining the structures of

25 Der_f I or Der_f II, producing peptides to be examined for their ability to influence B-cell and/or T-cell responses in house dust mite-sensitive individuals and selecting appropriate epitopes recognized by the cells. Synthetic amino acid

30 sequences which mimic those of the epitopes and which are capable of down regulating allergic response to Der_f I or Der_f II allergen can be

2091449

made. Proteins, peptides or antibodies of the present invention can also be used, in known methods, for detecting and diagnosing allergic response to Der_f I or Der_f II. For example, this
5 can be done by combining blood obtained from an individual to be assessed for sensitivity to one of these allergens with an isolated allergenic peptide of house dust mite, under conditions appropriate for binding of or stimulating components (e.g.,
10 antibodies, T cells, B cells) in the blood with the peptide and determining the extent to which such binding occurs. The Der_p I and Der_p II proteins or peptides can be used in a similar manner for desensitization and diagnosis of sensitivity. Der_f
15 and Der_p proteins or peptides can be administered together to treat an individual sensitive to both allergen types.

It is now also possible to design an agent or a drug capable of blocking or inhibiting the ability
20 of Der_f I or Der_f II to induce an allergic reaction in house dust mite-sensitive individuals. Such agents could be designed, for example, in such a manner that they would bind to relevant anti-Der_f I or anti-Der_f II IgEs, thus preventing IgE-allergen
25 binding and subsequent mast cell degranulation. Alternatively, such agents could bind to cellular components of the immune system, resulting in suppression or desensitization of the allergic response to these allergens. A non-restrictive
30 example of this is the use of appropriate B- and T-cell epitope peptides, or modifications thereof, based on the cDNA/protein structures of the present

2091449

invention to suppress the allergic response to Der_f I or Der_f II allergens. This can be carried out by defining the structures of B- and T-cell epitope peptides which affect B- and T-cell function in in vitro studies with blood cells from Der_f I or Der_f II-sensitive individuals. This can also be applied to Der_p I or Der_p II, in order to block allergic response to these allergens.

The cDNA encoding Der_f I or Der_f II or peptide including at least one epitope can be used to produce additional peptides, using known techniques such as gene cloning. A method of producing a protein or a peptide of the present invention can include, for example, culturing a host cell containing an expression vector which, in turn, contains DNA encoding all or a portion of a selected allergenic protein or peptide (e.g., Der_f I, Der_f II or a peptide including at least one epitope). Cells are cultured under conditions appropriate for expression of the DNA insert (production of the encoded protein or peptide). The expressed product is then recovered, using known techniques. Alternatively, the allergen or portion thereof can be synthesized using known mechanical or chemical techniques. As used herein, the term protein or peptide refers to proteins or peptides made by any of these techniques. The resulting peptide can, in turn, be used as described previously.

DNA to be used in any embodiment of this invention can be cDNA obtained as described herein or, alternatively, can be any oligodeoxynucleotide sequence having all or a portion of the sequence

2091449

represented in Figures 2 and 6, or their functional equivalent. Such oligodeoxynucleotide sequences can be produced chemically or mechanically, using known techniques. A functional equivalent of an

5 oligonucleotide sequence is one which is capable of hybridizing to a complementary oligonucleotide sequence to which the sequence (or corresponding sequence portions) of Figures 2 and 6 hybridizes and/or which encodes a product (e.g., a polypeptide

10 or peptide) having the same functional characteristics of the product encoded by the sequence (or corresponding sequence portion) represented in these figures. Whether a functional equivalent must meet one or both criteria will

15 depend on its use (e.g., if it is to be used only as an oligoprobe, it need meet only the first criterion and if it is to be used to produce house dust mite allergen, it need only meet the second criterion).

The structural information now available (e.g.,

20 DNA, protein/peptide sequences) can also be used to identify or define T cell epitope peptides and/or B cell epitope peptides which are of importance in allergic reactions to D. farinae allergens and to elucidate the mediators or mechanisms (e.g.,

25 interleukin-2, interleukin-4, gamma interferon) by which these reactions occur. This knowledge should make it possible to design peptide-based house dust mite therapeutic agents or drugs which can be used to modulate these responses.

30 The present invention will now be further illustrated by the following Examples, which are not intended to be limiting in any way.

2091449

EXAMPLESExample 1Isolation and Characterization of cDNA Coding for
Der f I5 MATERIALS AND METHODSDermatophagoides farinae culture

Mites were purchased from Commonwealth Serum Laboratories, Parkville, Australia.

Construction of the D. farinae cDNA λ gt11 library

- 10 Polyadenylated mRNA was isolated from live D. farinae mites and cDNA was synthesized by the RNase H method (Gubler, V. and B.J. Hoffman, Gene, 25:263-269 (1983)) using a kit (Amersham International, Bucks.). After the addition of EcoRI
- 15 linkers (New England Biolabs, Beverly, MA) the cDNA was ligated to alkaline phosphatase treated λ gt11 arms (Promega, Madison, WI). The ligated DNA was packaged and plated in E. coli Y1090 (r-) to produce a library of 2×10^4 recombinants.

20 Isolation of Der f I cDNA clones from the D. farinae cDNA λ gt11 library

Screening of the library was performed by hybridization with two probes comprising the two Der

2091449

p I cDNA BamHI fragments 1-348 and 349-857 generated by BamHI digestion of a derivative of the Der_p I cDNA which has had two BamHI restriction sites inserted between amino acid residues -1 and 1 and
5 between residues 116 and 117 by site-directed mutagenesis (Chua, K.Y. et al., J. Exp. Med., 167:175-182 (1988)). The probes were radiolabelled with ^{32}P by nick translation. Phage were plated at 20,000 pfu per 150mm petri dish and plaques were
10 lifted onto nitrocellulose (Schleicher and Schull, Dassel, FRG), denatured and baked (Maniatis, T. et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1982)).
Prehybridizations were performed for 2 hours at 42°C
15 in 50% formamide/5 x SSCE/1 x Denhardt's/poly C (0.1mg/ml)/poly U(0.1mg/ml) with hybridization overnight at 42°C at 10^6 cpm/ml. Post hybridization washes consisted of 15 min washes at room
temperature with 2 x sodium chloride citrate
20 (SSC)/0.1% sodium dodecylsulphate (SDS), 0.5 x SSC/0.1% SDS, 0.1 x SSC/0.1% SDS successively and a final wash at 50°C for 30 min in 0.1 x SSC/1% SDS.

Isolation of DNA from λ gt11 f 1 cDNA clones

Phage DNA from λ gt11 f 1 clones was prepared by
25 a rapid isolation procedure. Clarified phage plate lysate (1 ml) was mixed with 270 μL of 25% wt/vol polyethylene glycol (PEG 6000) in 2.5M NaCl and incubated at room temperature for 15 min. The mixture was then spun for 5 min in a microfuge
30 (Eppendorf, FRG), and the supernatant was removed. The pellet was dissolved in 100 μL of 10mM Tris/HCl

pH8.0 containing 1 mM EDTA and 100 mM NaCl (TE). This DNA preparation was extracted with phenol/TE, the phenol phase was washed with 100 μ l TE, the pooled aqueous phases were then extracted another 2
5 times with phenol/TE, 2 times with Leder phenol (phenol/chloroform/isoamylalcohol; 25:24:1), once with chloroform and the DNA was precipitated by ethanol.

DNA sequencing

10 To obtain clones for DNA sequence analysis, the λ gt11 fl phage DNA was digested with EcoRI restriction enzyme (Pharmacia, Uppsala, Sweden) and the DNA insert was ligated to EcoRI-digested M13-derived sequencing vectors mp18 and mp19
15 (Maniatis, T. et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1982)). Transformation was carried out using E. coli TG-1 and sequencing was performed by the dideoxynucleotide chain termination method (Sanger,
20 F. et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467 (1977)) using the SequenaseTM version 2.0 DNA sequencing kit (U.S.B., Cleveland, Ohio).

Polymerase chain reaction (PCR)

PCR was performed by the Taq DNA polymerase
25 method (Saiki, R.K. et al. Science, 239:487-491 (1988)) using the TaqPacTM kit (Biotech International, Bentley, WA) and the conditions recommended by the supplier with 10ng of target DNA and 10pmol of λ gt11 primers (New England BioLabs, Beverly, MA).

2091449

RESULTS

Isolation of Der_f I cDNA clones

Two clones expressing the major mite allergen Der_f I were isolated from the D. farinae cDNA λ gt11 library by their ability to hybridize with both of the Der_p I cDNA probes (nucleotides 1-348 and 349-857). This approach was adopted because amino acid sequencing had shown high homology (80%) between these two allergens (Thomas, W.R. et al., Advances in the Biosciences, 14:139-147 (1989)). Digestion of the λ gt11 fl clone DNA with EcoRI restriction enzyme to release the cDNA insert produced three Der_f I cDNA EcoRI fragments: one approximately 800 bases long and a doublet approximately 150 bases long. The Der_f I cDNA insert was also amplified from the phage DNA by the polymerase chain reaction (PCR) resulting in a PCR product of approximately 1.1-kb. Each Der_f I cDNA fragment was cloned separately into the M13-derived sequencing vectors mp18 and mp19 and sequenced.

DNA sequence analysis

The nucleotide sequence of the Der_f I cDNA was determined using the sequencing strategy shown in Figure 1. The complete sequence was shown to be 1084 bases long and included a 335-base long 5' proximal end sequence, a coding region for the entire native Der_f I protein of 223 amino acids with a derived molecular weight of 25,191 and an 80-base long 3' noncoding region (Fig. 2). The assignment of the threonine residue at position 1 as

2091449 - 30 -

the NH₂-terminal amino acid of Der_f I was based on data obtained by NH₂-terminal amino acid sequencing of the native protein and the predicted amino acid sequence of recombinant Der_p I (Chua, K.Y. et al., J. Exp. Med., 167:175-182 (1988)). The predicted amino acid sequence of the Der_f I cDNA in the NH₂-terminal region matched completely with that determined at the protein level (Figure 2).

The complete mature protein is coded by a single open reading frame terminating at the TGA stop codon at nucleotide position 1007-1009. The first ATG codon at nucleotide position 42-44 is presumed to be the translation initiation site since the subsequent sequence codes for a typical signal peptide sequence.

Amino Acid Sequence Analysis

The amino acid sequence predicted by nucleotide analysis is shown in Figure 2. As shown in the composite alignment of the amino acid sequence of mature Der_p I and Der_f I (Figure 3), high homology was observed between the two proteins. Sequence homology analysis revealed that the Der_f I protein showed 81% homology with the Der_p I protein as predicted by previous conventional amino acid sequencing. In particular, the residues making up the active site of Der_p I, based on those determined for papain, actinidin, cathepsin H, and cathepsin B, are also conserved in the Der_f I protein. The residues are glutamine (residue 29), glycine, serine and cysteine (residues 33-35), histidine (residue 171) and asparagine, serine and

tryptophan (residues 191-193) where the numbering refers to Der_f I. The predicted mature Der_f I amino acid sequence contains a potential N-glycosylation site (Asn-Thr-Ser) at position 53-55 which is also present as Asn-Gln-Ser at the equivalent position in Der_p I.

Analysis of the predicted amino acid sequence of the entire Der_f I cDNA insert has shown that, as for other cysteine proteases (Figure 4), the Der_f I protein has pre- and proform intermediates. As previously mentioned, the methionine residue at position -98 is presumed to be the initiation methionine. This assumption is based on the fact that firstly, the 5' proximal end sequence from residues -98 to -81 is composed predominantly of hydrophobic amino acid residues (72%), which is the characteristic feature of signal peptides (Von Heijne, G., EMBO J., 3:2315-2323 (1984)). Secondly, the lengths of the presumptive pre- (18 amino acid residues) and pro-peptides (80 residues) are similar to those for other cysteine proteases (Figure 4). Most cysteine proteases examined have about 120 preproenzyme residues (of which an average of 19 residues form the signal peptide) with cathepsin B the smallest with 80 (Ishidoh, K. et al., FEBS Letters, 226:32-37 (1987)). Der_f I falls within this range with a total of 98 preproenzyme residues.

By following the method for predicting signal-sequence cleavage sites outlined in Von Heijne, it is proposed that cleavage from the pre-Der_f I sequence for proenzyme formation occurs at the signal peptidase cleavage site lying between Ala

2091449

(-81) and Arg (-80) (Von Heijne, G., Eur. J. Biochem., 133:17-21 (1988) and J. Mol. Biol., 184:99-105 (1985)). Thus, the sequence from residues -98 to -81 codes for the leader peptide while the proenzyme moiety of Der_f I begins at residue Arg (-80) and ends at residue Glu (-1).

Example 2

Isolation and Characterization of cDNA Coding for Der_f II

10 MATERIALS AND METHODS

Amino acid sequence analysis

Preparation of λ gt11 *D. farinae* cDNA ligations

D. farinae was purchased from Commonwealth Serum Laboratories, Parkville, Australia, and used to prepare mRNA (polyadenylated RNA) as described (Stewart, G.A. and W.R. Thomas, Int. Arch. Allergy Appl Immunol., 83:384-389 (1987)). The mRNA was suspended at approximately 0.5 μ g/ μ l and 5 μ g used to prepare cDNA by the RNase H method (Gubler, U. and Hoffman, B.J., Gene, 25:263-269 (1983)) using a kit (Amersham International, Bucks). EcoRI linkers (Amersham, GGAATTCC) were attached according to the method described by Huynh et al., Constructing and screening cDNA libraries in gt10 and gt11, In: Glover, DNA Cloning vol. A practical approach pp. 47-78 IRL Press, Oxford (1985)). The DNA was then digested with EcoRI and recovered from an agarose gel purification by electrophoresis into a DEAE membrane (Schleicher and Schuell, Dassel, FRG,

2091449

NA-45) according to protocol 6.24 of Sambrook et al., (Sambrook et al., Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press (1989)), except 0.5M arginine base
5 was used for elution. The cDNA was then ligated in λ gt10 and λ gt11 at an arms to insert ratio of 2:1. Some was packaged for plaque libraries and an aliquot retained for isolating sequences by polymerase chain reaction as described below.

10 Isolation of Der f II cDNA by Polymerase Chain Reaction

To isolate Der f II cDNA, an oligonucleotide primer based on the N-terminal sequence of Der p II was made because their amino acid residues are
15 identical in these regions (Heymann, P.W. et al., J. Allergy Clin. Immunol., 83:1055-1087 (1989)). The primer GGATCCGATCAACTCGATGC-3' was used. The first GGATCC encodes a BamHI site and the following sequence GAT... encodes the first 4 residues of Der
20 p II. For the other primer the λ gt11 TTGACACCAGACC-AACTGGTAATG-3' reverse primer flanking the EcoRI cloning site was used (New England Biolabs, Beverly, MA). The Der p II primer was designed to have approximately 50-60% G-C and to end on the first or
25 second, rather than the third, base of a codon (Gould, S.J. et al., Proc. Natl. Acad. Sci., 86:1934-1938 (1989); Summer, R. and D. Tautz, Nucleic Acid Res., 17:6749 (1989)).

The PCR reactions were carried out in a final
30 reaction volume of 25 μ l containing 67mM Tris-HCL (pH8.8 at 25°C), 16.6mM $(\text{NH}_4)_2\text{SO}_4$, 40 μ M dNTPs, 5mM

2091449

- 34 -

2-mercaptoethanol, 6 μ M EDTA, 0.2mg/ml gelatin, 2mM
MgCl₂, 10pmoles of each primer and 2 units of Taq
polymerase. Approximately 0.001 μ g of target DNA was
added and the contents of the tube were mixed and
5 overlaid with paraffin oil. The tubes were
initially denatured at 95°C for 6 minutes, then
annealed at 55°C for 1 minute and extended at 72°C
for 2 minutes. Thereafter for 38 cycles, denaturing
was carried out for 30 seconds and annealing and
10 extension as before. In the final (40th) cycle, the
extension reaction was increased to 10 minutes to
ensure that all amplified products were full length.
The annealing temperature was deliberately set
slightly lower than the T_m of the oligonucleotide
15 primers (determined by the formula $T_m = 69.3 +$
 $0.41(G+C\%) - 650/\text{oligo length}$) to allow for mismatches
in the N-terminal primer.

5 μ l of the reaction was then checked for
amplified bands on a 1% agarose gel. The remainder
20 of the reaction mixture was extracted with chloro-
form to remove all of the paraffin oil and ethanol
precipitated prior to purification of the amplified
product on a low melting point agarose gel (Bio-Rad,
Richmond, CA).

25 Subcloning of PCR product

The ends of the purified PCR product were
filled in a reaction containing 10 mM Tris HCl, 10
mM MgCl₂, 50 mM NaCl, 0.025 mM dNTP and 1 μ l of
Klenow enzyme in a final volume of 100 μ l. The
30 reaction was carried out at 37°C for 15 minutes and
heat inactivated at 70°C for 10 minutes. The

2091449

mixture was Leder phenol extracted before ethanol precipitation. The resulting blunt ended DNA was ligated into M13mp18 digested with Sma I in a reaction containing 0.5M ATP, 1 X ligase buffer and 1 unit of T_4 ligase at 15°C for 24 hrs and transformed into E. coli TG1 made competent by the $CaCl_2$ method. The transformed cells were plated out as a lawn on L + G plates and grown overnight at 37°C.

10 Preparation of Single-stranded DNA Template for Sequencing

Isolated white plaques were picked using an orange stick into 2.5ml of an overnight culture of TG1 cells diluted 1 in 100 in 2 X TY broth, and grown at 37°C for 6 hours. The cultures were pelleted and the supernatant removed to a fresh tube. To a 1ml aliquot of this supernatant 270 μ l of 20% polyethylene glycol, 2.5M NaCl was added and the tube was vortexed before allowing it to stand at RT for 15 minutes. This was then spun down again and all traces of the supernatant were removed from the tube. The pellet was then resuspended in 100 μ l of 1 X TE buffer. At least 2 phenol:TE extractions were done, followed by 1 Leder phenol extraction and a $CHCl_3$ extraction. The DNA was precipitated in ethanol and resuspended in a final volume of 20 μ l of TE buffer.

DNA Analysis

DNA sequencing was performed with the dideoxy-nucleotide chain termination (Sanger, F. et al.).

2091449

Proc. Nat. Acad. Sci., 74:5463-5467 (1977)) using DNA produced from M13 derived vectors mp18 and mp19 in E. coli TG1 and T4 DNA polymerase (Sequenase version 2.0, USB Corp., Cleveland, Ohio; Restriction endonucleases were from Toyobo, (Osaka, Japan). All general procedures were by standard techniques (Sambrook, J. et al., A Laboratory Manual, 2d Ed. Cold Spring Harbor Laboratory Press (1989)). The sequence analysis was performed using the Mac Vector Software (IBI, New Haven, CT).

RESULTS

D. farinae cDNA ligated in λ gt11 was used to amplify a sequence using an oligonucleotide primer with homology to nucleotides coding for the 4 N terminal residues of Der p II and a reverse primer for the λ gt11 sequence flanking the coding site. Two major bands of about 500 bp and 300 bp were obtained when the product was gel electrophoresed. These were ligated into M13 mp18 and a number of clones containing the 500 bp fragment were analyzed by DNA sequencing. Three clones produced sequence data from the N-terminal primer end and one from the other orientation. Where the sequence data from the two directions overlapped, a complete match was found. One of the clones read from the N-terminal primer, contained a one-base deletion which shifted the reading frame. It was deduced to be a copying error, as the translated sequence from the other two clones matched the protein sequence for the first 20 amino acid residues of the allergen.

The sequence of the clones showing consensus and producing a correct reading frame is shown in Figure 6, along with the inferred amino acid sequence. It coded for a 129 residue protein with no N-glycosylation site and a calculated molecular weight of 14,021 kD. No homology was found when compared to other proteins on the GenBankTM data base (61.0 release). It did, however, show 88% amino acid residue homology with Der p II shown in the alignment in Figure 9. Seven out of the 16 changes were conservative. The conserved residues also include all the cysteines present at positions 8, 21, 27, 73 and 119. There was also considerable nucleotide homology, although the restriction enzyme map generated from the sequence data for commonly used enzymes is different from Der p II (Figure 8). The hydrophobicity plots of the translated sequence of Der f II and Der p II shown in Figure 10 are almost identical.

It is expected that there are sequence polymorphisms in the nucleic acid sequence coding for Der p I, Der p II, Der f I and Der f II, and it will be appreciated by one skilled in the art that one or more nucleotides in the nucleic acid sequences coding for Der p I, Der p II, Der f I and Der f II may vary among individual mites due to natural allelic variation. Any and all such nucleotide variations and resulting amino acid polymorphisms are within the scope of the invention. Amino acid polymorphisms were discovered during the sequencing of different Der p I clones. These polymorphisms are shown in Figure 11.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

Claims

1. An isolated nucleic acid encoding a *Der f* I protein allergen comprising the amino acid sequence shown in Figure 2, or a peptide comprising an epitope of said *Der f* I protein allergen, the epitope not being cross reactive with *Der p* I protein allergen.
2. An isolated nucleic acid encoding amino acid residues 1-223 of the amino acid sequence of *Der f* I shown in Figure 2.
3. An isolated nucleic acid encoding a *Der f* I protein allergen, or a peptide comprising an epitope of *Der f* I protein allergen, wherein the nucleic acid comprises the nucleotide sequence shown in Figure 2.
4. An isolated nucleic acid encoding a protein allergen of *Dermatophagoides*, wherein the nucleic acid comprises a nucleotide sequence which hybridizes under conditions of high stringency to the nucleotide sequence of *Der f* I shown in Figure 2 or the complement thereof, but does not hybridize under conditions of high stringency to the nucleotide sequence of *Der p* I.
5. An isolated protein allergen encoded by a nucleic acid as recited in any of claims 1-4.
6. An isolated *Der f* I protein allergen comprising the amino acid sequence shown in Figure 2, or a peptide comprising an epitope of *Der f* I protein allergen comprising the amino acid sequence shown in Figure 2, the epitope not being cross reactive with *Der p* I protein allergen.

7. An isolated *Der f* I protein allergen comprising amino acid residues I-223 of the amino acid sequence shown in Figure 2.
8. An isolated protein allergen or peptide as recited in any one of claims 5-7 produced recombinantly in a host cell transformed with a nucleic acid encoding the protein allergen or peptide or produced by chemical synthesis.
9. An isolated protein allergen or peptide as recited in any one of claims 5-7, wherein the epitope is a T or a B cell epitope.
10. An isolated protein allergen of *Dermatophagoides pteronyssinus*, *Der p* I, comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of *Der p* I (a) as represented in Figure 11; the amino acid sequence of *Der p* I (b) as represented in Figure 11; the amino acid sequence of *Der p* I (c) as represented in Figure 11; the amino acid sequence of *Der p* I (d) as represented in Figure 11; and the amino acid sequence of *Der p* I (e) as represented in Figure 11.
11. A composition comprising a protein allergen or peptide as recited in any one of claims 5-10 and a pharmaceutically acceptable carrier.
12. A composition comprising a protein allergen or peptide as recited in any one of claims 5-10 and at least one peptide derived from a protein allergen selected from *Der f* II, *Der p* I or *Der p* II, or any combination thereof, wherein each of said peptides comprises at least one T or B cell epitope.

13. The composition of claim 12, further comprising a pharmaceutically acceptable carrier.

14. Use of the protein allergen or peptide as recited in any one of claims 5-10, for the manufacture of a medicament for treating sensitivity to house dust mites.

15. Use of the protein allergen or peptide as recited in any one of claims 5-10, in combination with at least one peptide derived from a protein allergen selected from *Der f* II, *Der p* I, or *Der p* II, or any combination thereof, wherein each of the *Der p* I or *Der p* II peptides comprises at least one T or B cell epitope, for the manufacture of a medicament for the treatment of house dust mite sensitivity.

16. A method of detecting sensitivity in an individual to house dust mites, comprising combining a biological sample from the individual with the protein allergen or peptide as recited in any one of claims 5-10, under conditions appropriate for binding of components from the sample with the protein, and determining the extent to which such binding occurs.

17. A method of producing the protein allergen or peptide as recited in any one of claims 5-10, comprising culturing a host cell transformed with a nucleic acid encoding the protein allergen or peptide in medium to express the protein or peptide and isolating the protein or peptide from the medium.

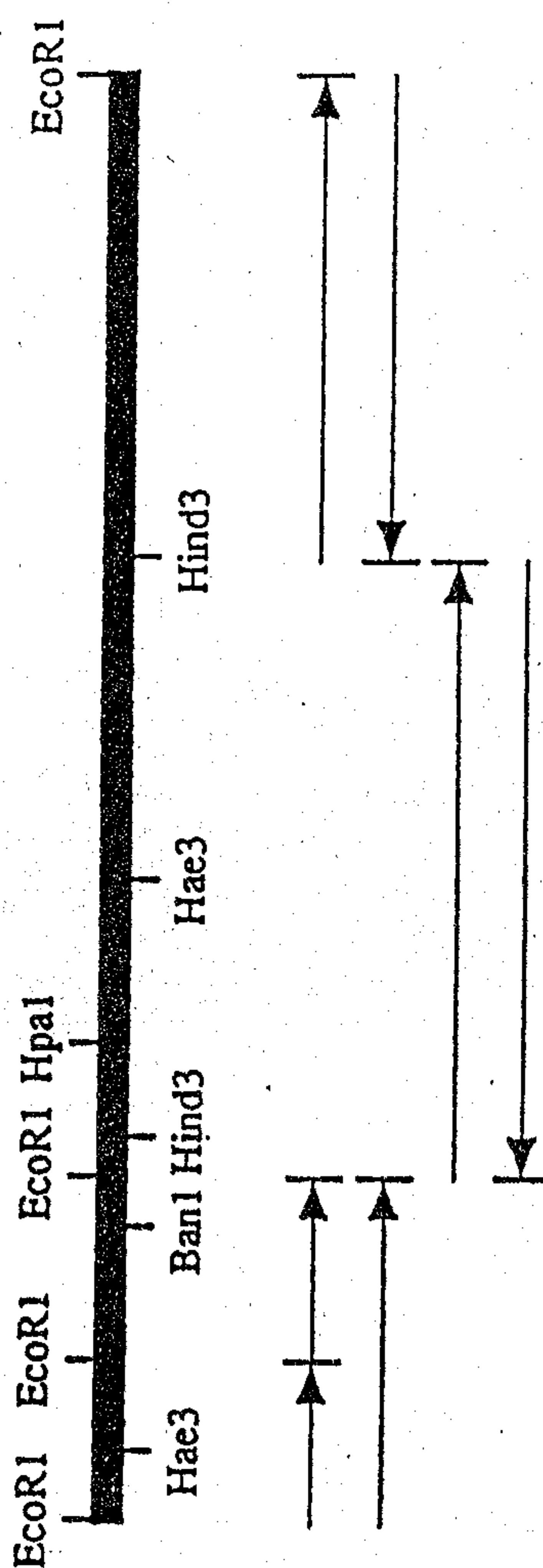


Fig. 1

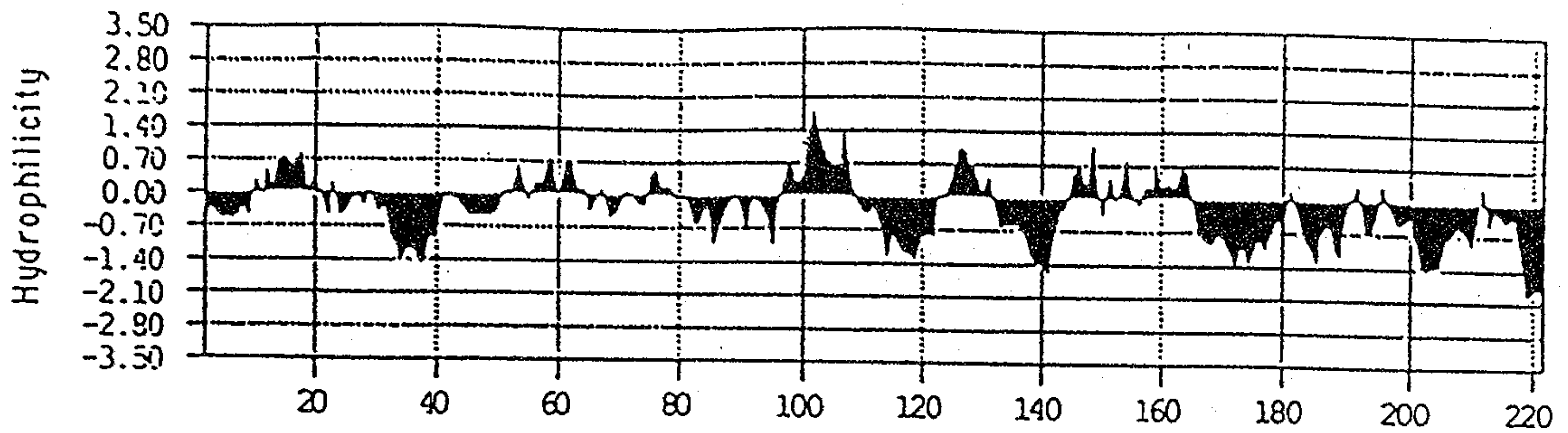
2/12

2091449

GAATTCCTGTTTCTTCCATCAAAATTAAAAATTCATCAAAA																	ATG	AAA	TTC	GTT	TTG	GCC	ATT	62
																	Met	Lys	Phe	Val	Leu	Ala	Ile	
-90																	-80							
GCC	TCT	TTG	TTG	GTA	TTG	AGC	ACT	GTT	TAT	GCT	CGT	CCA	GCT	TCA	ATC	AAA	ACT	116						
Ala	Ser	Leu	Leu	Val	Leu	Ser	Thr	Val	Tyr	Ala	Arg	Pro	Ala	Ser	Ile	Lys	Thr							
-70																	-60							
TTT	GAA	GAA	TTC	AAA	AAA	GCC	TTC	AAC	AAA	AAC	TAT	GCC	ACC	GTT	GAA	GAG	GAA	170						
Phe	Glu	Glu	Phe	Lys	Lys	Ala	Phe	Asn	Lys	Asn	Tyr	Ala	Thr	Val	Glu	Glu	Glu							
-50																	-40							
GAA	GTT	GCC	CGT	AAA	AAC	TTT	TTG	GAA	TCA	TTG	AAA	TAT	GTT	GAA	GCT	AAC	AAA	224						
Glu	Val	Ala	Arg	Lys	Asn	Phe	Leu	Glu	Ser	Leu	Lys	Tyr	Val	Glu	Ala	Asn	Lys							
-30																	-20							
GGT	GCC	ATC	AAC	CAT	TTG	TCC	GAT	TTG	TCA	TTG	GAT	GAA	TTC	AAA	AAC	CGT	TAT	278						
Gly	Ala	Ile	Asn	His	Leu	Ser	Asp	Leu	Ser	Leu	Asp	Glu	Phe	Lys	Asn	Arg	Tyr							
-10																	10							
TTG	ATG	AGT	GCT	GAA	GCT	TTT	GAA	CAA	CTC	AAA	ACT	CAA	TTC	GAT	TTG	AAT	GCC	332						
Leu	Met	Ser	Ala	Glu	Ala	Phe	Glu	Gln	Leu	Lys	Thr	Gln	Phe	Asp	Leu	Asn	Ala							
-1																	1							
GAA	ACA	AGC	GCT	TGC	CGT	ATC	AAT	TCG	GTT	AAC	GTT	CCA	TCG	GAA	TTG	GAT	TTA	386						
Glu	Thr	Ser	Ala	Cys	Arg	Ile	Asn	Ser	Val	Asn	Val	Pro	Ser	Glu	Leu	Asp	Leu							
20																	30							
CGA	TCA	CTG	CGA	ACT	GTC	ACT	CCA	ATC	CGT	ATG	CAA	GGA	GGC	TGT	GGT	TCA	TGT	440						
Arg	Ser	Leu	Arg	Thr	Val	Thr	Pro	Ile	Arg	Met	Gln	Gly	Gly	Cys	Gly	Ser	Cys							
40																	50							
TGG	GCT	TTC	TCT	GGT	GTT	GCC	GCA	ACT	GAA	TCA	GCT	TAT	TTG	GCC	TAC	CGT	AAC	494						
Trp	Ala	Phe	Ser	Gly	Val	Ala	Ala	Thr	Glu	Ser	Ala	Tyr	Leu	Ala	Tyr	Arg	Asn							
60																	70							
ACG	TCT	TTG	GAT	CTT	TCT	GAA	CAG	GAA	CTC	GTC	GAT	TGC	GCA	TCT	CAA	CAC	GGA	548						
Thr	Ser	Leu	Asp	Leu	Ser	Glu	Gln	Glu	Leu	Val	Asp	Cys	Ala	Ser	Gln	His	Gly							
80																	90							
TGT	CAC	GGC	GAT	ACA	ATA	CCA	AGA	GGC	ATC	GAA	TAC	ATC	CAA	CAA	AAT	GGT	GTC	602						
Cys	His	Gly	Asp	Thr	Ile	Pro	Arg	Gly	Ile	Glu	Tyr	Ile	Gln	Gln	Asn	Gly	Val							
100																	110							
GTT	GAA	GAA	AGA	AGC	TAT	CCA	TAC	GTT	GCA	CGA	GAA	CAA	CGA	TGC	CGA	CGA	CCA	656						
Val	Glu	Glu	Arg	Ser	Tyr	Pro	Tyr	Val	Ala	Arg	Glu	Gln	Arg	Cys	Arg	Arg	Pro							
120																	130							
AAT	TCG	CAA	CAT	TAC	GGT	ATC	TCA	AAC	TAC	TGC	CAA	ATT	TAT	CCA	CCA	GAT	GTG	710						
Asn	Ser	Gln	His	Tyr	Gly	Ile	Ser	Asn	Tyr	Cys	Gln	Ile	Tyr	Pro	Pro	Asp	Val							
140																	150							
AAA	CAA	ATC	CGT	GAA	GCT	TTG	ACT	CAA	ACA	CAC	ACA	GCT	ATT	GCC	GTC	ATT	ATT	764						
Lys	Gln	Ile	Arg	Glu	Ala	Leu	Thr	Gln	Thr	His	Thr	Ala	Ile	Ala	Val	Ile	Ile							
160																	170							
GGC	ATC	AAA	GAT	TTG	AGA	GCT	TTC	CAA	CAT	TAT	GAT	GGA	CGA	ACA	ATC	ATT	CAA	818						
Gly	Ile	Lys	Asp	Leu	Arg	Ala	Phe	Gln	His	Tyr	Asp	Gly	Arg	Thr	Ile	Ile	Gln							
180																	190							
CAT	GAC	AAT	GGT	TAT	CAA	CCA	AAC	TAT	CAT	GCC	GTC	AAC	ATT	GTC	GGT	TAC	GGA	872						
His	Asp	Asn	Gly	Tyr	Gln	Pro	Asn	Tyr	His	Ala	Val	Asn	Ile	Val	Gly	Tyr	Gly							
200																	210							
AGT	ACA	CAA	GGC	GAC	GAT	TAT	TGG	ATC	GTA	CGA	AAC	AGT	TGG	GAT	ACT	ACC	TGG	926						
Ser	Thr	Gln	Gly	Asp	Asp	Tyr	Trp	Ile	Val	Arg	Asn	Ser	Trp	Asp	Thr	Thr	Trp							
220																	223							
GGA	GAT	AGC	GGA	TAC	GGA	TAT	TTC	CAA	GCC	GGA	AAC	AAC	CTC	ATG	ATG	ATC	GAA	980						
Gly	Asp	Ser	Gly	Tyr	Gly	Tyr	Phe	Gln	Ala	Gly	Asn	Asn	Leu	Met	Met	Ile	Glu							
220																	223							
CAA	TAT	CCA	TAT	GTT	GTA	ATC	ATG	TGA	ACATTTGAAATTGAATATATTTATTTGTTTTCAAAAT									1044						
Gln	Tyr	Pro	Tyr	Val	Val	Ile	Met																	
220																	223							
220	223																							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
220																	223							
2																								

5/12 2091449

Der o1



Der f1

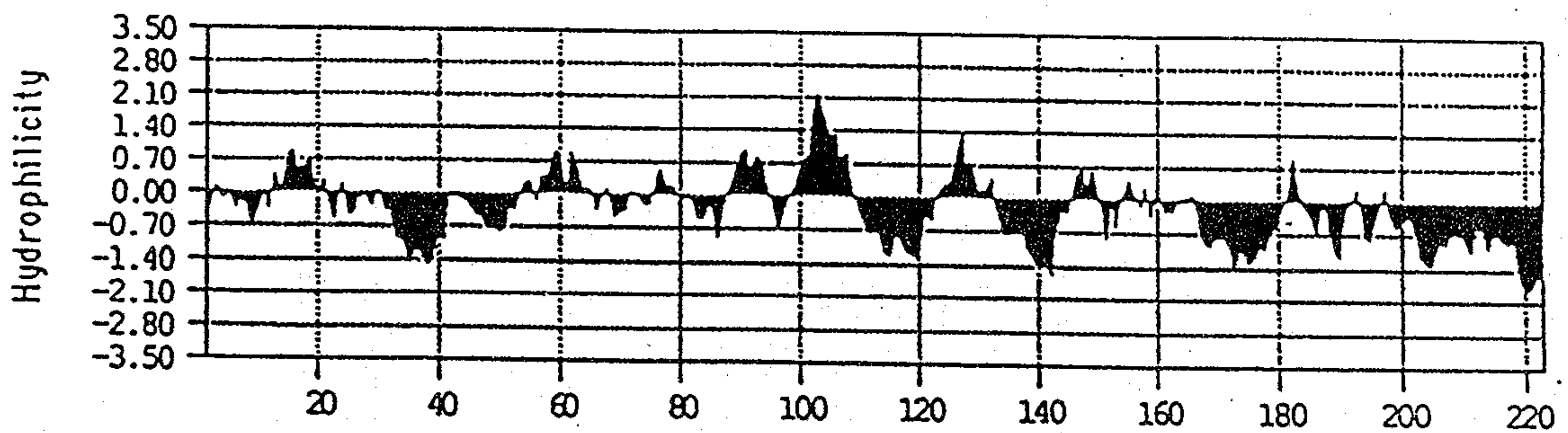


Fig. 5

SUBSTITUTE SHEET

6/12 2091449

																	10																		
GAT	CAA	GTC	GAT	GTT	AAA	GAT	TGT	GCC	AAC	AAT	GAA	ATC	AAA	AAA	GTA	ATG	51																		
Asp	Gln	Val	Asp	Val	Lys	Asp	Cys	Ala	Asn	Asn	Glu	Ile	Lys	Lys	Val	Met																			
																	20																		
GTC	GAT	GGT	TGC	CAT	GGT	TCT	GAT	CCA	TGC	ATC	ATC	CAT	CGT	GGT	AAA	CCA	102																		
Val	Asp	Gly	Cys	His	Gly	Ser	Asp	Pro	Cys	Ile	Ile	His	Arg	Gly	Lys	Pro																			
																	40																		
TTC	ACT	TTG	GAA	GCC	TTA	TTC	GAT	GCC	AAC	CAA	AAC	ACT	AAA	ACC	GCT	AAA	153																		
Phe	Thr	Leu	Glu	Ala	Leu	Phe	Asp	Ala	Asn	Gln	Asn	Thr	Lys	Thr	Ala	Lys																			
																	60																		
ACT	GAA	ATC	AAA	GCC	AGC	CTC	GAT	GGT	CTT	GAA	ATT	GAT	GTT	CCC	GGT	ATT	204																		
Thr	Glu	Ile	Lys	Ala	Ser	Leu	Asp	Gly	Leu	Glu	Ile	Asp	Val	Pro	Gly	Ile																			
																	70																		
GAT	ACC	AAT	GCT	TGC	CAT	TTT	ATG	AAA	TGT	CCA	TTG	GTT	AAA	GGT	CAA	CAA	255																		
Asp	Thr	Asn	Ala	Cys	His	Phe	Met	Lys	Cys	Pro	Leu	Val	Lys	Gly	Gln	Gln																			
																	90																		
TAT	GAT	GCC	AAA	TAT	ACA	TGG	AAT	GTG	CCG	AAA	ATT	GCA	CCA	AAA	TCT	GAA	306																		
Tyr	Asp	Ala	Lys	Tyr	Thr	Trp	Asn	Val	Pro	Lys	Ile	Ala	Pro	Lys	Ser	Glu																			
																	110																		
AAC	GTT	GTC	GTT	ACA	GTC	AAA	CTT	GTT	GGT	GAT	AAT	GGT	GTT	TTG	GCT	TGC	357																		
Asn	Val	Val	Val	Thr	Val	Lys	Leu	Val	Gly	Asp	Asn	Gly	Val	Leu	Ala	Cys																			
																	120																		
GCT	ATT	GCT	ACC	CAC	GCT	AAA	ATC	CGT	GAT	TAAAAAAAAAAAAATAAATATGAAAATT											414														
Ala	Ile	Ala	Thr	His	Ala	Lys	Ile	Arg	Asp																										
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		
																	129																		

7/12 2091449

-23												-10															
AAA	AAC	CGA	TTT	TTG	ATG	AGT	GCA	GAA	GCT	TTT	GAA	CAC	CTC	AAA	ACT	48											
Lys	Asn	Arg	Phe	Leu	Met	Ser	Ala	Glu	Ala	Phe	Glu	His	Leu	Lys	Thr												
						-1	1																				
CAA	TTC	GAT	TTG	AAT	GCT	GAA	ACT	AAC	GCC	TGC	AGT	ATC	AAT	GGA	AAT	96											
Gln	Phe	Asp	Leu	Asn	Ala	Glu	Thr	Asn	Ala	Cys	Ser	Ile	Asn	Gly	Asn												
10										20																	
GCT	CCA	GCT	GAA	ATC	GAT	TTG	CGA	CAA	ATG	CGA	ACT	GTC	ACT	CCC	ATT	144											
Ala	Pro	Ala	Glu	Ile	Asp	Leu	Arg	Gln	Met	Arg	Thr	Val	Thr	Pro	Ile												
				30										40													
CGT	ATG	CAA	GGA	GGC	TGT	GGT	TCA	TGT	TGG	GCT	TTC	TCT	GGT	GTT	GCC	192											
Arg	Met	Gln	Gly	Gly	Cys	Gly	Ser	Cys	Trp	Ala	Phe	Ser	Gly	Val	Ala												
						50																					
GCA	ACT	GAA	TCA	GCT	TAT	TTG	GCT	CAC	CGT	AAT	CAA	TCA	TTG	GAT	CTT	240											
Ala	Thr	Glu	Ser	Ala	Tyr	Leu	Ala	His	Arg	Asn	Gln	Ser	Leu	Asp	Leu												
		60										70															
GCT	GAA	CAA	GAA	TTA	GTC	GAT	TGT	GCT	TCC	CAA	CAC	GGT	TGT	CAT	GGT	288											
Ala	Glu	Gln	Glu	Leu	Val	Asp	Cys	Ala	Ser	Gln	His	Gly	Cys	His	Gly												
						80																					
GAT	ACC	ATT	CCA	CGT	GGT	ATT	GAA	TAC	ATC	CAA	CAT	AAT	GGT	GTC	GTC	336											
Asp	Thr	Ile	Pro	Arg	Gly	Ile	Glu	Tyr	Ile	Gln	His	Asn	Gly	Val	Val												
90										100																	
CAA	GAA	AGC	TAC	TAT	CGA	TAC	GTT	GCA	CGA	GAA	CAA	TCA	TGC	CGA	CGA	384											
Gln	Glu	Ser	Tyr	Tyr	Arg	Tyr	Val	Ala	Arg	Glu	Gln	Ser	Cys	Arg	Arg												
				110										120													
CCA	AAT	GCA	CAA	CGT	TTC	GGT	ATC	TCA	AAC	TAT	TGC	CAA	ATT	TAC	CCA	432											
Pro	Asn	Ala	Gln	Arg	Phe	Gly	Ile	Ser	Asn	Tyr	Cys	Gln	Ile	Tyr	Pro												
						130																					
CCA	AAT	GCA	AAC	AAA	ATT	CGT	GAA	GCT	TTG	GCT	CAA	ACC	CAC	AGC	GCT	480											
Pro	Asn	Ala	Asn	Lys	Ile	Arg	Glu	Ala	Leu	Ala	Gln	Thr	His	Ser	Ala												
		140										150															
ATT	GCC	GTC	ATT	ATT	GGC	ATC	AAA	GAT	TTA	GAC	GCA	TTC	CGT	CAT	TAT	528											
Ile	Ala	Val	Ile	Ile	Gly	Ile	Lys	Asp	Leu	Asp	Ala	Phe	Arg	His	Tyr												
					160																						
GAT	GGC	CGA	ACA	ATC	ATT	CAA	CGC	GAT	AAT	GGT	TAC	CAA	CCA	AAC	TAT	576											
Asp	Gly	Arg	Thr	Ile	Ile	Gln	Arg	Asp	Asn	Gly	Tyr	Gln	Pro	Asn	Tyr												
170										180																	
CAC	GCT	GTC	AAC	ATT	GTT	GGT	TAC	AGT	AAC	GCA	CAA	GGT	GTC	GAT	TAT	624											
His	Ala	Val	Asn	Ile	Val	Gly	Tyr	Ser	Asn	Ala	Gln	Gly	Val	Asp	Tyr												
				190										200													
TGG	ATC	GTA	CGA	AAC	AGT	TGG	GAT	ACC	AAT	TGG	GGT	GAT	AAT	GGT	TAC	672											
Trp	Ile	Val	Arg	Asn	Ser	Trp	Asp	Thr	Asn	Trp	Gly	Asp	Asn	Gly	Tyr												
						210																					
GGT	TAT	TTT	GCT	GCC	AAC	ATC	GAT	TTG	ATG	ATG	ATT	GAA	GAA	TAT	CCA	720											
Gly	Tyr	Phe	Ala	Ala	Asn	Ile	Asp	Leu	Met	Met	Ile	Glu	Glu	Tyr	Pro												
		220		222																							
TAT	GTT	GTC	ATT	CTC	TAAACAAAAAGACAATTTCTTATATGATTGTCACATAATTTATT										778												
Tyr	Val	Val	Ile	Leu																							
TAAATCAAAATTTTGTAGAAAATGAATAAATTCATTCACAAAAATTAAAAA																841											
AAAAAAAAAAAAAAAAAAAA																857											

Fig. 7

8/12 2091449

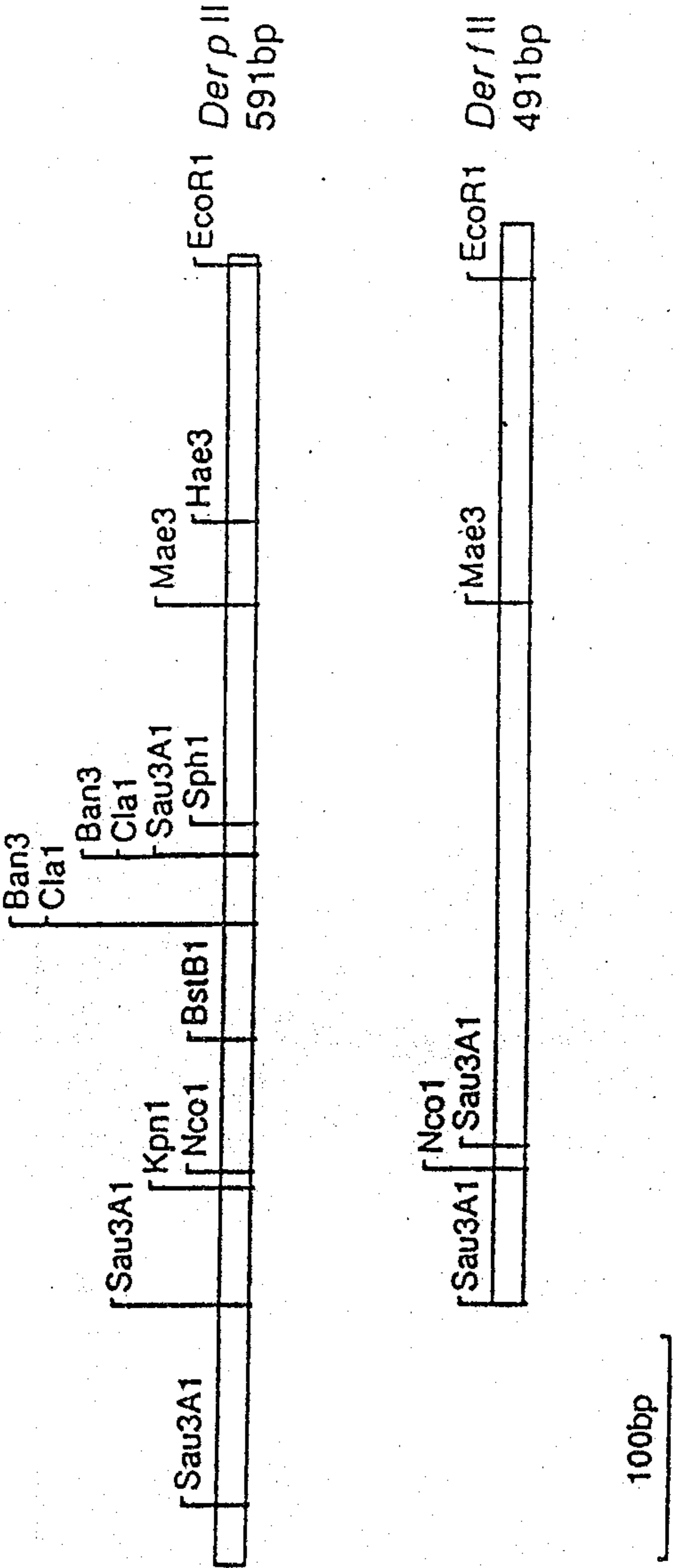


Fig. 8

9/12 2091449

Dp II - CACAAATTCTTCTTTCTTCCTTACTACTGATCATTAATCTGAAAACAAAA -50

-16 -10
M Y K I L C L S L L

Dp II - CCAAACAAACCATTCAAAATGATGTACAAAATTTGTGTCTTTCATTGTT -100

-1 1 10
V A A V A R D Q V D V K D C A N

Dp II - GGTCGCAGCCGTTGCTCGTGATCAAGTCGATGTCAAAGATTGTGCCAATC -150

Df II - GATCAAGTCGATGTTAAAGATTGTGCCAACA -31

20

Dp II - ATGAAATCAAAAAAGTTTGGTACCAGGATGCCATGGTTCAGAACCATGT -200

Df II - ATGAAATCAAAAAAGTAATGGTTCGATGGTTGCCATGGTTCTGATCCATGC -81

N M D D

30 40

Dp II - ATCATTCATCGTGGTAAACCATTCCAATTGGAAGCCGTTTTCGAAGCCAA -250

Df II - ATCATCCATCGTGGTAAACCATTCACTTTGGAAGCCTTATTGATGCCAA -131

T L D

50 60

Dp II - CCAAAACACAAAAACGGCTAAAATTGAAATCAAAGCCTCAATCGATGGTT -300

Df II - CCAAAACACTAAAACCGCTAAAATGAAATCAAAGCCAGCCTCGATGGTC -181

T L

70

Dp II - TAGAAGTTGATGTTCCCGGTATCGATCCAAATGCATGCCATTACATGAAA -350

Df II - TTGAAATTGATGTTCCCGGTATTGATACCAATGCTTGCCATTTTATGAAA -231

I T F

80 90

Dp II - TGCCCATTTGGTTAAAGGACAACAATATGATATTAAATATACATGGAATGT -400

Df II - TGTCCATTGGTTAAAGGTCAACAATATGATGCCAAATATACATGGAATGT -281

A

100 110

Dp II - TCCGAAAATTGCACCAAAATCTGAAAATGTTGTCGTCCTGTTAAAGTTA -450

Df II - GCCGAAAATTGCACCAAAATCTGAAAACGTTGTCGTTACAGTCAAACCTTG -331

L

Fig. 9

SHEET 1 of 2

SUBSTITUTE SHEET

11/12

2091449

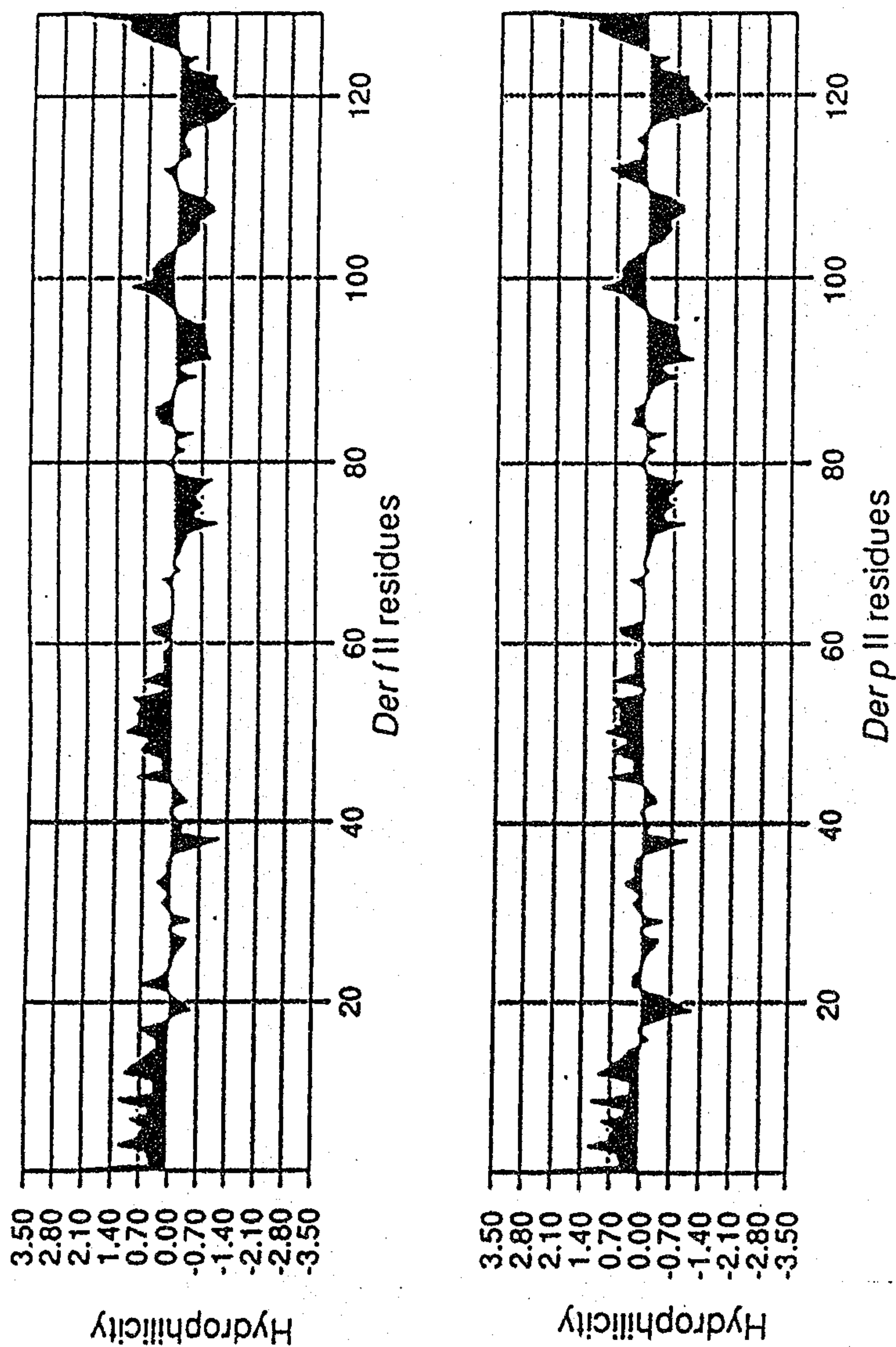


Fig. 10

SUBSTITUTE SHEET

12/12

2091449

Der p I SEQUENCES

	10	20	30	40	50	60
Der p I (a)	TNAC	SINGNA	PAEIDL	QMR	TVTPIR	MOGG
Der p I (b)	CGSC	WAFSGV	AATESA	YLAH	RNQSLD	LAEQ
Der p I (c)	---	---	---	---	---	---
Der p I (d)	---	---	---	---	---	---
Der p I (a)	ELVD	CASQHG	CHGDTI	PRGI	EYIQHNGV	VQ
Der p I (b)	---	---	---	---	---	---
Der p I (c)	---	---	---	---	---	---
Der p I (d)	---	---	---	---	---	---
Der p I (a)	PPNANKI	REA	LAQTHSA	IAV	IIGIKDLDAF	
Der p I (b)	---	---	---	---	---	---
Der p I (c)	---	---	---	---	---	---
Der p I (d)	---	---	---	---	---	---
Der p I (e)	---	---	---	---	---	---
Der p I (a)	QGV	DYWI	VRN	SWDTN	WGDNG	YGYFAANIDL
Der p I (b)	---	---	---	---	---	---
Der p I (c)	---	---	---	---	---	---
Der p I (d)	---	---	---	---	---	---
Der p I (e)	---	---	---	---	---	---

SUBSTITUTE SHEET

Fig. 11