House dust mites cause heavy atopic diseases such as asthma and dermatitis. Among allergens from *Dermatophagoides farinae*, Der f 2 shows the highest positive rate for atopic patients, but its biological function in mites has been perfectly unknown, as well as the functions of its homologs in human and other animals. We have determined the tertiary structure of Der f 2 by multidimensional nuclear magnetic resonance spectroscopy. Der f 2 was found to be a single-domain protein of immunoglobulin fold, and its structure was the most similar to those of the two regulatory domains of transglutaminase. This fact, binding to the bacterial surface, and other small pieces of information hinted that Der f 2 is related to the innate antibacterial defense system in mites. The immunoglobulin E epitopes are also discussed on the basis of the tertiary structure.

Mites are the closest animals to human life and their relation is inseparable in the modern residential environment. House dust mites cause heavy atopic diseases such as asthma and dermatitis, which are rapidly increasing worldwide, especially among the children in developed countries. *Dermatophagoides farinae* and *D. pteronyssinus* are recognized as the main sources of house dust allergens. Among their allergens, the group-2 allergen proteins, Der f 2 and Der p 2, show the highest positive rate for atopic patients (1) so they are called major allergens. Their sequences are 88% identical (2, 3), and their cross-reactivity was well confirmed (1). They are also homologous to the major allergen from the mite *Lepidoglyphus destructor* (4), which is important in farming environments. These proteins are 125–129 amino acid residues long and have three intramolecular disulfide bonds.

Although the properties of these proteins related to their allergenicity have been well characterized, their biological function in mites is unknown. Homologous proteins were also found in human epididymis, cow milk, and moth trachea (5–7) and named HE1, EPV20 and esr16, respectively, the first two found in human epididymis, cow milk, and moth trachea (5–7) and named HE1, EPV20 and esr16, respectively, the first two of which are glycoproteins unlike the group-2 mite allergens. However, there have been no clues for their functions except for their expression patterns. This is in contrast to the case of the other major allergen group including Der f 1 and Der p 1, because they were found to be cysteine proteases, and their activity was suggested to be involved in the induction of allergic responses (8). Therefore, the innate functions of Der f 2 and Der p 2 have interested many allergy researchers.

It is generally accepted that allergic symptoms are initiated by the specific binding of allergens to immunoglobulin E (IgE) antibodies, which cross-link the high affinity IgE receptors on mast cells and basophils (9). So monovalent ligands to allergen-specific IgE are expected to block IgE receptor aggregation. Structure determination of allergens will offer the basis of design of such drugs. The tertiary structures of some pollen allergens have been previously reported including ragweed allergens Amb t 5 and Amb a 5, a major birch allergen Bet v 1, and a minor birch allergen profilin (10–14). However, the tertiary structure of Der f 2 and the following drug-design processes are exceptionally urgent considering the serious mental influences on atopic children.

In this study, we have determined the tertiary structure of Der f 2 by multidimensional nuclear magnetic resonance (NMR) spectroscopy. Unexpectedly we found that Der f 2 is a single-domain protein of immunoglobulin fold. There are few single-domain proteins of this fold, and this protein is soluble and monomeric. We feel it is of interest also in terms of the evolution of protein folds. The structural similarity to the two regulatory domains of transglutaminase and other small pieces of information prompted us to suppose that Der f 2 is a component of the innate antibacterial defense system in mites. Furthermore, we found that Der f 2 binds to the surface of bacteria, which is the first clue to the biological function of this class of proteins. We also discuss previous work related to the immunoglobulin E epitopes on the basis of the tertiary structure.

**EXPERIMENTAL PROCEDURES**

Sample Preparation—Escherichia coli BL21 cells harboring pFLT1 (pGEMEX1 derivative containing cDNA for clone 1 of Der f 2; see Ref. 15) were cultured in M9 minimum medium in the presence of [15N]ammonium chloride. Labeled recombinant Der f 2 was expressed and purified as described (16). NMR samples contained 1.5 mM of Der f 2 in 90% H2O, 10% D2O, 140 mM N-Octyl-β-D-glucoside, 0.01% NaN3 at pH 5.6.

NMR Spectroscopy—NMR spectra were acquired at 55 °C on a Varian Unityplus 600 NMR spectrometer equipped with a triple resonance pulse field gradient probe. The sequential assignment of the 1H, 13C and 15N resonances were made as described and assigned (16).

### References

1. The abbreviations used are: IgE, immunoglobulin E; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; RMSD, root mean square difference.
RESULTS AND DISCUSSION

Assignments and Structure Determination—Backbone sequential assignments for Der f 2 (129 amino acid residues) were obtained by a strategy using a combination of four triple resonance measurements, 3D HN(CO)CA, HNCA, CBCA(CO)NH, CBCANH, and were complete except for H² and N of Asp-1 and Thr-123, and C6 of Glu-53. The nitrogen chemical shifts of prolines were not assigned. Side-chain assignments were obtained principally by 3D C(CO)NH, HC/C/H-TOCSY and HC/C/H-COSY, and partially extended using NOE data. All nonexchangeable resonances were assigned except for Lys-33 C6 and Lys-126 H2, H6, and H4, whose assignments were not fixed owing to heavy overlappings. In addition, H2, H4, and Nδ of Trp-92 were not detected, possibly because of fast proton solvent exchange enhanced by interaction with added detergent molecules.

The secondary structure was determined as all β, using tertiary NOEs between backbone protons (Fig. 1), which had been estimated beforehand by the results of chemical shift index method (20). Using 1086 distance constraints extracted after assignments of 3D 15N- and 13C-edited NOE spectroscopy spectra, we obtained 10 structures from 20 calculations. A summary of the structural statistics for a set of the final structures and for the mean structure is presented in Table I. There were no violations above 0.6 Å in any of the structures, and the number of violations above 0.3 Å ranged from 4 to 13. The deviation from ideal bond lengths was 0.004 Å. These figures are relatively good, considering that our distance constraint set was tighter than those in the typical three-level classification. However, since the number of NOE constraints was not large, the quality of the structure should be regarded as medium.

Description and Evaluation of the Structure—Fig. 2 shows the tertiary structure determination of Der f 2. The root mean square difference (RMSD) value of the 10 final structures from the minimized averaged structure was 0.90 ± 0.15 Å for the backbone (N, Cα, C′) atoms of residues 1–129 and 1.44 ± 0.17 Å for the nonhydrogen atoms of the same residue range. We also calculated local RMSD values for the backbone atoms (N,
C, C') at each residue after the best fit superposition of the whole backbone (data not shown), and found that the residues of large local RMSD value were distributed around the whole sequence. Therefore the reason the structural convergence was not excellent is not due to local flexibility but to the number of NOE distance constraints.

Der f 2 has an immunoglobulin fold (Fig. 2b). The topology corresponds to s-type (21). One β-sheet consists of three strands of residues 13–19 (a), 36–43 (b), and 85–92 (e), and the other sheet is composed of four β-strands 62–65 (c'), 49–57 (c), 104–111 (f), and 116–127 (g), and an additional short strand of residues 5–8 (tentatively named as a'). They include three β-bulges at the positions 14–15, 117–118, and 123–124. Der f 2 has three disulfide bonds at the positions 8–119, 21–27, and 73–78 (22), but no disulfide bridge between the β-strands b and f, which is often found in immunoglobulin-fold domains.

The immunoglobulin fold is the most ubiquitous module and is distributed among many protein superfamilies of different functions (21). There are, however, few single-domain proteins with the immunoglobulin fold. Der f 2 is not membrane bound like Thy-1, a subunit in a protein complex like 2-microglobulin, nor does it polymerize into filaments like the major sperm protein from nematode. Therefore, this simple immunoglobulin, Der f 2, has been recognized as unique to the mite allergen cysteine proteases (8). The fact that Der f 2 inhibited guinea pig liver transglutaminase is also interesting as it relates to the structural similarity described above.

Implications for Innate Functions in Mites—Although the biological function of Der f 2 is unknown, the above findings reminded us of the innate immune system of invertebrates (28). In invertebrates that do not have immunoglobulins, the coagulation system is prominent as an antimicrobial response. Two types of coagulation mechanisms have been reported, each of which are associated with transglutaminase and a cascade of serine proteases, respectively. Der f 3, a minor mite allergen, is a serine protease that has a similar substrate specificity to blood coagulation factor XII and is reported to activate the human serine protease cascade (29).

We found that Der f 2 binds the surface of E. coli cells (Fig. 4) in a manner similar to hemolin, a bacteria-binding protein from moths composed of four immunoglobulin-like domains (30). Der f 2 has a cluster of nine basic amino acids (Fig. 3a), which implies a negatively charged target surface. Preliminary results show that Der f 2 does not bind to strains K12 or BL21. All of the three types of mite allergens mentioned are localized in the gastrointestinal tract, mouth region and feces (31, 32), and the feces, which are suggested to cause allergic symptoms.

Then we aligned the primary sequences of these factor XIII domains (26) to those of Der f 2 (2) and homologous proteins (3–7) on the basis of tertiary structures (Fig. 3b). Although the sequences in each group are highly variable, we could find identical amino acid residues shared by both groups at many positions. The numbers of such residues were 32 both in the third and fourth domains of factor XIII, which is much larger than the number in the second domain of vascular cell adhesion molecule, which is 22. Phe-41, which is located at the center of domains, is perfectly conserved among the two groups shown in Fig. 3b. In particular, the factor XIII fourth domain is well aligned to the group-2 mite allergens without large insertions or deletions except for three disulfide bond-related segments.

Intriguingly, Der f 1, the other major mite allergen, is a cysteine protease, and this protease family has been shown to be evolutionally related to the second domain of factor XIII (27).

We found that Der f 2 binds the surface of E. coli cells (Fig. 4) in a manner similar to hemolin, a bacteria-binding protein from moths composed of four immunoglobulin-like domains (30). Der f 2 has a cluster of nine basic amino acids (Fig. 3a), which implies a negatively charged target surface. Preliminary results show that Der f 2 does not bind to strains K12 or BL21. All of the three types of mite allergens mentioned are localized in the gastrointestinal tract, mouth region and feces (31, 32), and the feces, which are suggested to cause allergic symptoms.

2 S. Kojima, unpublished observation.
(33), have microbial degradation activities (34). HE1, EPV20, and esr16, the Der f 2 homologs from human, cow, and moth, respectively, are included in epithelial mucosae, where antibacterial proteins are excreted (35–37). These observations imply that Der f 2 is a component of the antibacterial defense system in mites. A survey of recent literature data (38–40) made us realize that Bet v 1, conalbumin and lactoferrin, which show the highest positive rate for sera from allergic patients of birch pollen, hen egg, and cow milk, respectively, are components of antimicrobial host defense systems.

**IgE Epitopes**—Since the cloning of the group-2 allergens, IgE and T-cell epitopes have been intensely studied. The experiments using 14 synthetic peptides of 15 residues in length spanning the entire sequence of Der p 2 showed that IgE antisera do not bind to most peptides; the peptide comprising residues 65–78 bound IgE, but its activity was extremely weak (41). Therefore recognition by IgE depends strongly on the conformation of Der p 2, and considering the intense homology, it is probably also the case for Der f 2. Truncation of N- or C-terminal short sequences (42) or destruction of the disulfide bond 8–119 (43) reduced IgE-binding activity severely, which corresponds to our result that Der f 2 is composed of only one domain.

Nishiyama et al. (42) made site-directed mutants at residues 1–21, 70–81, and 114–129 of Der f 2, which were selected considering the studies mentioned above, and measured their bacterial proteins are excreted (35–37). These observations imply that Der f 2 is a component of the antibacterial defense system in mites. A survey of recent literature data (38–40) made us realize that Bet v 1, conalbumin and lactoferrin, which show the highest positive rate for sera from allergic patients of birch pollen, hen egg, and cow milk, respectively, are components of antimicrobial host defense systems.

**IgE Epitopes**—Since the cloning of the group-2 allergens, IgE and T-cell epitopes have been intensely studied. The experiments using 14 synthetic peptides of 15 residues in length spanning the entire sequence of Der p 2 showed that IgE antisera do not bind to most peptides; the peptide comprising residues 65–78 bound IgE, but its activity was extremely weak (41). Therefore recognition by IgE depends strongly on the conformation of Der p 2, and considering the intense homology, it is probably also the case for Der f 2. Truncation of N- or C-terminal short sequences (42) or destruction of the disulfide bond 8–119 (43) reduced IgE-binding activity severely, which corresponds to our result that Der f 2 is composed of only one domain.

Nishiyama et al. (42) made site-directed mutants at residues 1–21, 70–81, and 114–129 of Der f 2, which were selected considering the studies mentioned above, and measured their bacterial proteins are excreted (35–37). These observations imply that Der f 2 is a component of the antibacterial defense system in mites. A survey of recent literature data (38–40) made us realize that Bet v 1, conalbumin and lactoferrin, which show the highest positive rate for sera from allergic patients of birch pollen, hen egg, and cow milk, respectively, are components of antimicrobial host defense systems.

**IgE Epitopes**—Since the cloning of the group-2 allergens, IgE and T-cell epitopes have been intensely studied. The experiments using 14 synthetic peptides of 15 residues in length spanning the entire sequence of Der p 2 showed that IgE antisera do not bind to most peptides; the peptide comprising residues 65–78 bound IgE, but its activity was extremely weak (41). Therefore recognition by IgE depends strongly on the conformation of Der p 2, and considering the intense homology, it is probably also the case for Der f 2. Truncation of N- or C-terminal short sequences (42) or destruction of the disulfide bond 8–119 (43) reduced IgE-binding activity severely, which corresponds to our result that Der f 2 is composed of only one domain.

Nishiyama et al. (42) made site-directed mutants at residues 1–21, 70–81, and 114–129 of Der f 2, which were selected considering the studies mentioned above, and measured their bacterial proteins are excreted (35–37). These observations imply that Der f 2 is a component of the antibacterial defense system in mites. A survey of recent literature data (38–40) made us realize that Bet v 1, conalbumin and lactoferrin, which show the highest positive rate for sera from allergic patients of birch pollen, hen egg, and cow milk, respectively, are components of antimicrobial host defense systems.
IgE-binding activities. Using their results, we mapped the molecular surface for those substitutions that decreased IgE binding (Fig. 5). This figure suggests that two IgE epitope areas on the surface. One epitope area includes Asp-7, Asn-10 and Lys-18, and the second one includes Cys-73, Phe-75, Lys-77, and Cys-78. However, the borders of these areas are distorted at the residues Asp-19 and Asn-71. Experiments that can judge whether each decrease of IgE binding is due to global destabilization or local effects on the allergen-antibody interface might improve epitope definition. Although additional amino acid substitutions are desired to clarify the borders and judge the existence of other epitope areas, our structure suggests mosaic distribution of IgE epitope areas and provides strategies for their complete characterization.

The engineered Der f 2 has already begun to be applied for immunotherapy strategies. For example, the mutated allergen in which the disulfide bond 8–119 was disrupted retained complete activity to stimulate T-cell proliferation (43). However, developments of monovalent ligands to allergen-specific receptor aggregation. Since the protein HE1 is a human homologue of Der f 2, chimeric proteins of HE1, and Der f 2 would be effective for stimulating discussions and advice.

Acknowledgements—We thank Dr. Mark B. Swindells of Helix Research Institute, Dr. Daisuke Kohda of Biomolecular Engineering Research Institute, Prof. Hikoichi Sakai of Japan Women’s University, Dr. Hajime Karasuyama of Tokyo Metropolitan Institute of Medical Science, Dr. Koji Nagata of Tokyo University, and our colleagues in the laboratories of Tokyo Metropolitan Institute of Medical Science and Asahi Breweries, Ltd. for stimulating discussions and advice.