THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 32, Issue of August 9, pp. 19243–19250, 1996

Printed in U.S.A.

Secondary Structure and Tertiary Fold of the Birch Pollen Allergen Bet v 1 in Solution*

(Received for publication, March 18, 1996, and in revised form, April 25, 1996)

Cornelius Faber‡, Almut Lindemann‡, Heinrich Sticht‡, Andrzej Ejchart‡, Andreas Kungl§, Markus Susani¶, Rainer W. Frank‖, Dietrich Kraft, Michael Breitenbach‡‡, and Paul Rösch‡§§**

*From ‡Lehrstuhl für Biopolymere, Universität Bayreuth, Universitätsstr. 30, D-95447 Bayreuth, Germany, §Institut für Genetik und Allgemeine Biologie, Universität Salzburg, A-5020 Salzburg, Austria, ‖‡‡Zentrum für Molekulare Biologie Heidelberg (ZMBH), INF 282, D-69120 Heidelberg, Germany, ¶Advanced Biological Systems, Billrothstr. 11, A-5020 Salzburg, Austria, §Sandoz Research Institute, Brunnerstr. 59, A-1235 Wien, Austria, and the **Institut für Allgemeine und Experimentelle Pathologie, Währinger Gürtel 18–20, A-1090 Wien, Austria*

Bet v 1 is the major birch pollen allergen and therefore the main cause of type I allergies observed in early spring. It is composed of 159 amino acid residues adding up to a molecular mass of 17 kDa. We determined the secondary structure and tertiary fold of full-length Bet v 1 by NMR spectroscopy. Two- and three-dimensional NMR measurements suggest that Bet v 1 is a globular monomer in solution with a high content of well defined secondary structure. Of the total of 159 residues, 135 could be sequentially assigned, using an improved assignment strategy based mainly on heteronuclear experiments. An improved strategy for structure calculation revealed three helices and two β-sheets as major elements of secondary structure. The globular tertiary structure is mainly stabilized by two antiparallel β-sheets. The two helices at the C terminus are in accordance with the results from the solution structure of the chemically synthesized peptide Bet v 1-(125–154). This peptide is composed of two helices connected by a hinge. The structural features of Bet v 1 are highly similar to those found in the *Ambrosia* allergen Amb t V.

In Europe and North America about 20% of the population suffer from type I (IgE-mediated) allergies. Typical manifestations of this hypersensitivity are hay fever, dermatitis, asthma, and, in extreme cases, anaphylactic shock. During a first contact with an allergenic substance, genetically predisposed individuals are sensitized. In these individuals, a decisive but not well understood step in the development of the disease is the class switch to produce IgE at relatively high levels. In a later stage of the disease contact with allergen leads to cross-linking of IgE receptors on mast cells, inducing the release of histamine and other mediators, which causes the immediate allergic reaction. The main source of allergens in the moderate climate zones of the earth are proteins from pollen, insects, foodstuff, animal epithelia, and moulds. The most important pollens are of grasses, weeds, and early flowering trees. Especially those trees within the order *Fagales* possess potent allergens. The most widespread one among those is Bet v 1, the main allergen of the birch (*Betula verrucosa*). Bet v 1 is composed of 159 amino acid residues and has a molecular mass of 17 kDa. It was first cloned and sequenced by Breiteneder *et al.* (1). It shows sequence similarity to the pollen allergens of hazel (2), hornbeam (3), and alder (4).

Determination of the spatial structure of Bet v 1 may, in comparison with structures of other allergens, lead to the identification of B cell epitopes. The only structure of an allergenic protein solved so far is that of Amb t V, the allergen from *Ambrosia* (ragweed) pollen (5). Amb t V contains a C-terminal helix as the major T cell epitope. This finding is in accordance with a theory proposed by Margalit *et al.* (6) stating that amphipathic α-helices are the preferred conformations of T cell epitopes. It was possible to distinguish seven T cell epitopes on Bet v 1 (7). Two of these are situated near the C terminus (positions 145–156 and 147–158). Several secondary structure prediction algorithms suggested helical conformation for residues 124–154 of Bet v 1 (8). The high conservation of this sequence region in the allergens of hazel and alder supports the assumption that a C-terminal helix containing a T cell epitope is a common structural feature of allergens.

In this study we used information obtained from two-dimensional ^{1}H-^{1}H, ^{1}H-^{15}N and three-dimensional ^{1}H-^{1}H-^{15}N NMR spectroscopy with uniformly ^{15}N-labeled Bet v 1. Taking into account the special role of the predicted C-terminal helix as a T cell epitope we also decided to analyze the structure and stability of the isolated, synthetic Bet v 1-(125–154) peptide.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification of Recombinant Bet v 1—Bet v 1 from overexpressing *Escherichia coli* was purified according to the published purification protocol (9), but under nondenaturating conditions, and stored as lyophilized product. In total, 25 mg of recombinant protein were available. ^{15}N-Labeled Bet v 1 was purified from overexpressing *E. coli* grown on minimal medium (1 liter of culture contained 17.1 g of $Na_2HPO_4 \cdot 12H_2O$, 3 g of KH_2HPO_4, 0.5 g of NaCl, 1 g of NH_4Cl, 8 g of glucose, 2 ml of 1 M $MgSO_4$, 0.1 ml of 1 M $CaCl_2$, 2 ml of 10 mM Fe(III) citrate, 50 mg of ampicillin, 2 ml of trace element solution (100 mg $ZnSO_4 \cdot 7H_2O$, 30 mg of $MnCl_2 \cdot 4H_2O$, 300 mg of H_3BO_3, 200 mg of $CoCl_2 \cdot 6H_2O$, 20 mg of $NiCl_2 \cdot 6H_2O$, 900 mg of $Na_2MoO_4 \cdot 2H_2O$, 20 mg of Na_2SeO_3 in 1000 ml H_2O)). For labeling with ^{15}N, ammonium chloride was replaced by the respective ^{15}N-enriched salt (Campro Scientific, Emmerich, Germany). A total of 30 mg of labeled Bet v 1 was available for the spectroscopic studies.

Bet v 1-(125–154) was synthesized on an automated continuous flow peptide synthesizer on polystyrene-polyethylene glycol (Tentagel) using standard Fmoc (9-fluorenylmethyloxycarbonyl) protocols and PyBop (benzotriazol-1-yl)-*N*-oxytripyrrolidinophosphonium hexafluorophosphate as activator (10). The product was purified by reversed phase high performance liquid chromatography to a purity of more than 95%. Mass spectrometry (Shimadzu Compact MALDI), Edman degradation,

* This work was supported by grants from the Fonds der Chemischen Industrie (to H. S. and P. R.) and the Deutsche Forschungsgemeinschaft (to P. R.) and by Austrian Academy of Science Apart Fellowship 246 (to A. K.).The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§§ To whom correspondence should be addressed: Lehrstuhl für Biopolymere, Universität Bayreuth, Universitätsstr. 30, D-95447 Bayreuth, Germany. Tel.: 49 921 553540; Fax: 49 921 553544; E-mail: paul.roesch@uni-bayreuth.de.

and amino acid analysis showed correct mass, amino acid composition, and amino acid sequence. The synthetic peptide was freeze-dried and stored at −20 °C.

NMR Sample Preparation—Two different samples of Bet v 1 were prepared: a 2 mM solution of non-isotope-labeled Bet v 1 in 50 mM sodium phosphate, 5 mM NaCl, pH 6.3, and a 4 mM solution of uniformly ^{15}N-labeled Bet v 1 in 50 mM sodium phosphate, pH 6.5.

Bet v 1-(125–154) was dissolved at a final concentration of 4 mM in a mixture of H_2O/deuterated TFE (F_3C-CD_2OH) (4:6), containing 20 mM sodium phosphate, pH 6.3.

NMR Spectroscopy—All NMR experiments were carried out on a Bruker AMX 600 spectrometer at 308 K. The homonuclear double quantum-filtered COSY, clean TOCSY,[1] and NOESY were performed in H_2O/D_2O (9:1) and followed standard methods (11, 12). NOESY experiments were recorded with mixing times of 100 and 200 ms in H_2O/D_2O (9:1) and in 99.994% D_2O. Clean TOCSY experiments were recorded with mixing times of 70, 80, and 90 ms. 4092 × 1024 data points were collected with a spectral width of 7264 Hz in both dimensions. A sine-bell-squared filter with a phase shift of $\pi/3$ or $\pi/4$ prior to Fourier transformation was used. Application of zero-filling resulted in a data size of 4092 × 2048 data points in the frequency domain.

Two-dimensional ^{15}N-^{1}H experiments were performed in H_2O/D_2O (9:1). HMQC experiments (13, 14) were performed, collecting 4092 × 512 data points with a spectral width of 7246 Hz in F_2 dimension and 3000 Hz and 7298 Hz in F_1 dimension. A relayed HMQC-NOESY experiment (15, 16) was recorded with a mixing time of 200 ms. 4092 × 512 data points were collected with a spectral width of 7246 Hz in the F_2 dimension and 3000 Hz in the F_1 dimension. For a relayed HMQC-TOCSY experiment (15, 16) 4092 × 472 data points were collected with a spectral width of 7246 Hz in the F_2- and 3000 Hz in the F_1 dimension. A sinebell-squared filter with a phase shift of $\pi/4$ prior to Fourier transformation was used. Single zero-filling in the F_1 dimension was used for the data from the NOESY experiment.

Three-dimensional ^{1}H-^{15}N-^{1}H TOCSY-HMQC (17) and NOESY-HMQC experiments (17–20) were recorded collecting 2048 × 128 × 280 (300 for NOESY) data points. The mixing time was set to 200 ms for the NOESY and 60 ms for the TOCSY experiment. The spectral width was 7246 Hz in the F_1 and the F_3 dimension and 2635 Hz in the F_2 dimension. Prior to Fourier transformation a sinebell-squared filter with a phase shift of $\pi/2$ was applied. After zero filling, Fourier transformation, and discarding the half of each spectrum that contains no ^{15}N-^{1}H correlations, a set of 1024 × 128 × 256 data points resulted for the NOESY spectrum and a set of 2048 × 128 × 256 data points for the TOCSY spectrum.

All spectra were acquired in the phase-sensitive mode with quadrature detection in all dimensions using the time-proportional phase incrementation technique (21). The solvent signal was suppressed by continuous coherent irradiation prior to the first excitation pulse and during the mixing time of the NOESY experiments. Base-line correction up to 7th order was performed for all two-dimensional spectra along the F_2 dimension and for the three-dimensional spectra along the F_3 dimension (22). In addition to the standard Bruker spectrometer control software, the NDEE software package[2] was used for data processing on X-window workstations. Chemical shift values are reported relative to 2,2-dimethyl-2-silapentane-5-sulfonate for ^{1}H and to liquid NH_3 for ^{15}N chemical shifts, respectively (23).

Distance information was obtained from two- and three-dimensional NOESY spectra in H_2O/D_2O (9:1) with mixing times of 200 and 100 ms. NOESY cross-peak intensities were grouped into three distance categories according to their relative intensities: strong, 0.18–0.27 nm; medium, 0.18–0.33 nm; weak, 0.18–0.5 nm (24). Peak intensities were estimated from the number of contours in contour plots of NOESY spectra. For distance limits involving nonstereospecifically assigned protons, the standard pseudoatom corrections were added (25). Taking into account the apparent higher intensity of methyl resonances, 0.05 nm was added to distances involving methyl groups (26). In the case of flipping aromatic rings, up to 0.2 nm was added to affected distances (27).

Hydrogen bonding restraints were obtained from the interpretation of NOESY spectra that were collected after 20 h in D_2O. A slow amide proton exchange was taken as indication that the proton was involved in a hydrogen bond (12).

Molecular Dynamics Calculations—The X-PLOR 3.1 package (28) was used to calculate three-dimensional structures. For the structure determination of the peptide the standard distance geometry, simulated annealing and refinement protocols were used (28). The only modification was an increase of the simulation time in the refinement to 30 ps and the explicit treatment of the van der Waals and electrostatic interactions.

The global fold of the full-length Bet v 1 was determined by the application of a new "random" simulated annealing protocol that was kindly provided by M. Nilges (EMBL, Heidelberg). All NOEs were specified directly in terms of the proton chemical shift assignments, treating frequency-degenerated NOEs as ambiguous (29).

The simulated annealing protocol was similar to the random simulated annealing protocol described in Ref. 29 with some minor modifications. For initial structure calculations all atomic coordinates were chosen randomly in a 20-Å cube. The nonbonded interactions were calculated only between a subset of atoms including C^α atoms and one side-chain atom using van der Waals radii of 2.25-Å and a nonbonded cut-off of 12-Å. 360 ps of molecular dynamics were calculated at 2000 K, while successively increasing the force constant for bond lengths from 0.05 to 10 kcal/(mol·Å^2) and for bond angles from 0.05 to 10 kcal/(mol·rad). Velocities were reassigned if a temperature of 2200 K was exceeded. For the NOE effective energy term a soft square well potential function was applied, and a weak harmonic potential (K = 0.0001 kcal/(mol·Å^2)) was used to prevent the atoms from escaping too far from the origin. At this stage of the calculation, 1158 interresidual ambiguous NOEs were used as distance restraints. Due to the frequency degeneration of the NOEs, 8354 assignments were consistent with the experimental data corresponding to an average of 7.2 possible assignments per NOE.

For the regularization of the derived structures the "dgsa" protocol was used, which is supplied with the X-PLOR 3.1 program package (28). 40 ps of molecular dynamics were calculated at 2000 K followed by a slow cooling to 100 K in 60 ps. A time step of 3 fs was applied throughout the protocol.

Of the resulting 50 structures, 20 structures were selected that showed the lowest energy and the least violation of the experimental data.

In order to obtain unambiguous assignments, all NOEs were checked against the selected structures. If an NOE could be attributed to the same proton pair in more than 50% of the structures, the NOE was considered as assigned with certainty.

Analysis of the Secondary Structure—Elements of secondary structure were deduced by the inspection of the pattern of the unambiguous NOEs, chemical shift analysis (30), and slow amide proton exchange.

In addition, the calculated structures were checked for the existence of secondary structural elements by DSSP analysis (31). Hydrogen bonds were deduced from N–O distances of less than 3.5 Å (32).

The resulting elements of secondary structure were compared with the results of several secondary structure prediction algorithms obtained by the mail server of the Institute de Biologie et Chimie des Proteines (Lyon).

For the identification of homologies on the level of the secondary structure the TOPITS algorithm (33) was applied, which is provided by the PredictProtein Server (34) at the EMBL (Heidelberg).

RESULTS AND DISCUSSION

Two-dimensional NMR spectra of Bet v 1 show good dispersion, and resonances occur between −0.2 and 9.7 ppm. For the amide protons the resonances stretch from 6.1 to 9.7 ppm. In spite of this, the resonances are overlapping strongly due to the size of the 159-amino acid protein. Thus, the standard techniques (12) did not allow a resonance assignment for more than a few residues, making isotope labeling necessary.

The two-dimensional ^{15}N-^{1}H relayed HMQC-TOCSY spectrum of Bet v 1 shows 176 N-H resonances. 20 of these originate from amide groups of the side chains of the asparagines and glutamines (Fig. 1). They were easily recognized for two reasons. First, two resonances occurred always pairwise at exactly the same ^{15}N shift. Second, the resonances were accompanied by shoulders systematically displaced by 0.6 ppm to higher

[1] The abbreviations used are: TOCSY, total correlation spectroscopy; clean TOCSY, TOCSY with suppression of NOESY-type cross-peaks; DSSP, definition of secondary structure of proteins; HMQC, heteronuclear multiple quantum coherence spectroscopy; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; RMSD, root mean square deviation; TFE, trifluoroethanol.

[2] F. Herrmann, and P. Rösch, unpublished; the program is available on request.

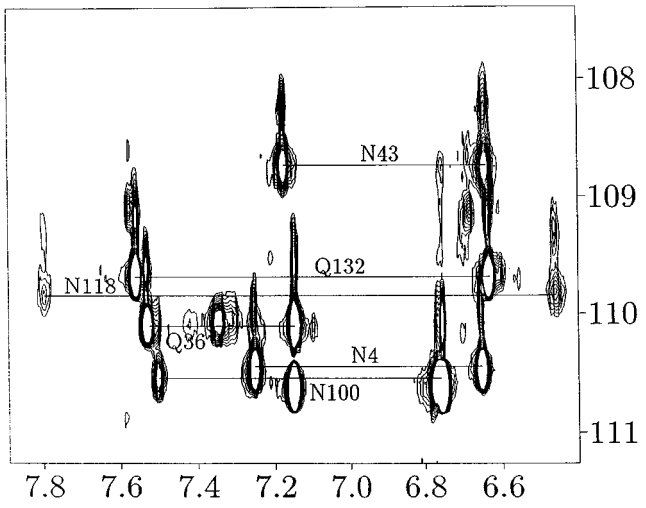

FIG. 1. **^{15}N-^{1}H relayed HMQC-TOCSY spectrum of Bet v 1.** Typical resonances of amide groups of asparagines and glutamines are shown.

field along the ^{15}N axis. These shoulders originate from N^1H^2H groups in which the chemical shielding of the nitrogen nuclei was affected by isotope exchange with the solvent (35). An additional three resonances were attributed to the NH of the arginine side chain. They were found at the typical chemical shift of about 85 ppm along the ^{15}N axis. Another resonance was found at 176.22 ppm and attributed to an NH of a histidine ring. With the additional information from the three-dimensional HMQC-TOCSY spectrum we were able to distinguish 149 spin systems in the homonuclear two-dimensional spectra. The identification and sequential assignment of the spin systems proved to be a very complicated and cumbersome task, because many spin systems showed untypical chemical shifts or missing resonances. The problem of strong degeneration of chemical shifts could be overcome by the analysis of the three-dimensional-spectra. In addition to the standard assignment strategy described in (17, 36) we used the two-dimensional ^{1}H-^{15}N NOESY spectrum for assignment purposes.

The strategy for the chain tracing in a ^{1}H-^{15}N relayed HMQC-NOESY spectrum is explained with Fig. 2 as an example. In addition to the intraresidual N_iH_i resonance, two interresidual resonances each can be observed at the same ^{1}H as well as at the same ^{15}N chemical shift. Each resonance along each axis originates from one of the two sequential neighbors. The two intraresidual resonances $N_{i-1}H_{i-1}$ and N_iH_i form a rectangle with the two corresponding interresidual resonances. Thus, the intraresidual N_iH_i resonance can be determined from the intraresidual $N_{i-1}H_{i-1}$ resonance and both interresidual resonances. Starting from the intraresidual $N_{i-1}H_{i-1}$ resonance (following the *solid arrow* in Fig. 2) the interresidual $N_{i-1}H_i$ resonance can be found at the same ^{15}N chemical shift but at a different ^{1}H chemical shift. Keeping this very ^{1}H chemical shift, the interresidual N_iH_{i+1} resonance can be found at a different ^{15}N chemical shift. Now keeping this ^{15}N chemical shift, the interresidual $N_{i+1}H_{i+2}$ resonance can be located at a different ^{1}H chemical shift. An alternative path for the chain tracing follows the *dashed arrows*. With both interresidual resonances for each pair of sequential neighbors, the intraresidual resonances can easily be determined. Chain tracing by this strategy can be performed much faster than in three-dimensional spectra and at a better resolution. On the other hand, it depends on the existence of NH-NH resonances and is therefore limited to helical sections or to availability of further information. In any case it supplements and confirms the information that can be obtained with standard strategies.

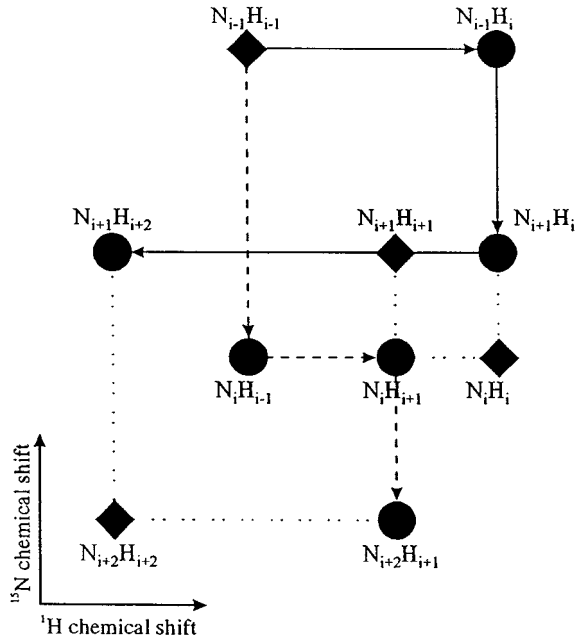

FIG. 2. **Schematic depiction of chain tracing in a ^{1}H-^{15}N-NOESY spectrum.** Axes are as in Fig. 1. ♦ represents an N_jH_j resonance; ● represents a sequential $N_jH_{j\pm1}$ resonance. *Solid arrows* indicate the path of chain tracing as explained in the text; *dashed arrows* indicate the alternative path.

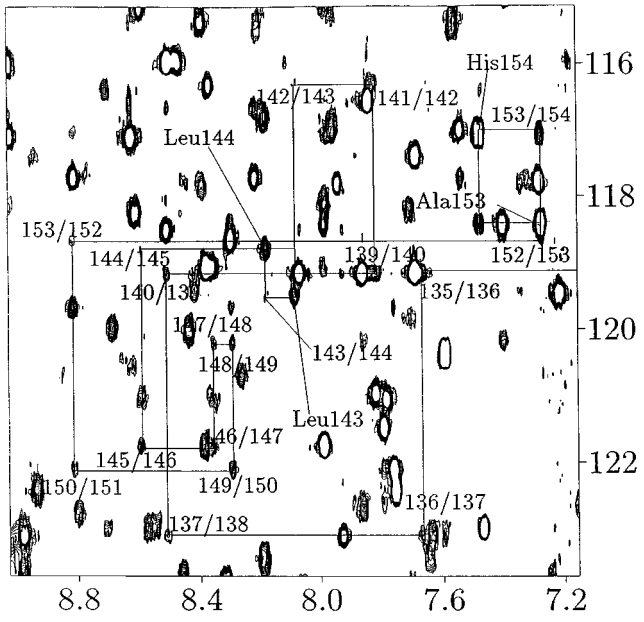

FIG. 3. **200-ms ^{15}N-^{1}H relayed HMQC-NOESY spectrum of Bet v 1.** Chain tracing is shown for residues 154–133 as explained in the text. For His^{154} and Ala^{153} both interresidual NOEs are shown. The ^{1}H chemical shift of Glu^{142} and Gly^{141} is degenerated. For Leu^{143} and Leu^{144} the second interresidual NOE is only observed in lower plot layers.

For Bet v 1 the sequential resonance assignment could solely be based on NOEs between backbone amide protons for residues Ala^{15}–Phe^{30}, Gln^{132}–Gly^{140}, and Leu^{144}–His^{154}, suggesting helical conformation in these regions. In Fig. 3, chain tracing starts with His^{154}. Both interresidual NOEs with Ala^{153} are shown. The chain tracing was possible up to Glu^{141}, which is frequency-degenerated in the ^{1}H dimension with Gly^{140}. Additional information from three-dimensional spectra removed the ambiguity. Further chain tracing from Gly^{140} to Ala^{135} is also shown.

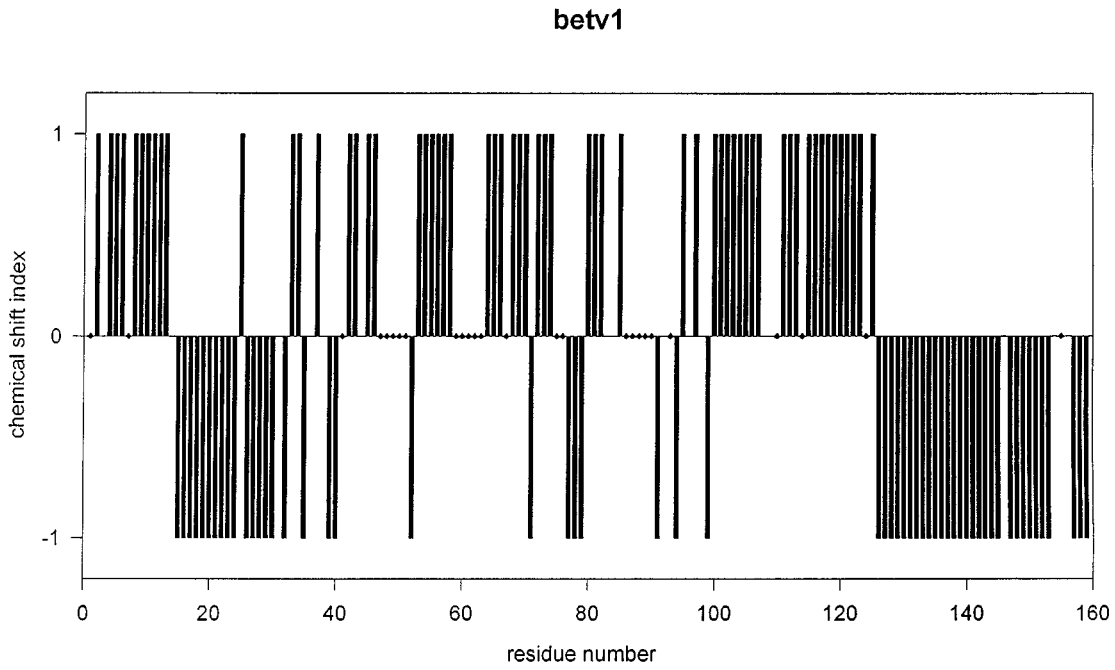

FIG. 4. **Chemical shift index diagram according to Wishart *et al.* (30).** Chemical shift index value 1 represents C^{α} proton resonances that are shifted more than 0.1 ppm downfield as compared with random coil values, and index value −1 represents an upfield shift by more than 0.1 ppm. Random coil values were obtained from the improved data set from Wishart *et al.* (23). ♦ indicates that the chemical shift of the C^{α} proton could not be determined. A high content of secondary structure elements is suggested by the pattern of the shifts.

For other spin systems the resonances of all protons were needed for the sequential assignment. In the case of missing NH-NH resonances, the origin of NOEs could not be attributed with certainty from the three-dimensional spectra. Here the two-dimensional spectra allowed removal of most ambiguities because of the better resolution. Many spin systems showed resonances far from the expected values for a peptide in random coil conformation, or resonances were missing completely for some spin systems. Nevertheless, 135 of the total 159 amino acid residues could finally be assigned sequentially, including stretches of 40 amino acids at either end of the protein almost without interruption. The only sections longer than two residues that could not be assigned were the glycine- and proline-rich regions Asn^{47}–Gly^{51}, and Pro^{59}–Pro^{63} as well as Ile^{86}–Pro^{90}.

We used these chemical shift data to perform a secondary structure estimate according to the chemical shift index strategy (23, 30). This procedure is based on the correlation between chemical shifts of C^{α} proton resonances of consecutive amino acids and local secondary structure. C^{α} proton resonances that are shifted to high field by more than 0.1 ppm compared with the corresponding random coil value for more than three consecutive amino acids indicate local α-helical structure. C^{α} proton resonances, shifted to low field for three or more consecutive residues correspond to a β-sheet. For Bet v 1, the pattern of the shifts (Fig. 4) suggests a high content of secondary structure elements including a long helix near the C terminus, as already deduced above. Nevertheless, a succession of two shorter helices connected by a turn cannot be ruled out from this data.

A further helix is suggested between residues Ala^{15} and Phe^{30}. The extreme downfield shift of the C^{α} proton resonance of Asp^{25} (5.35 ppm) does not necessarily indicate an interruption of the helix. The proximity of the aromatic ring of Phe^{22}, especially in an α-helical conformation, may lead to ring current effects that are likely to cause extreme deviations from the random coil values. This assumption is corroborated by the strong upfield shift of the C^{α} proton resonances of Gly^{26}, which

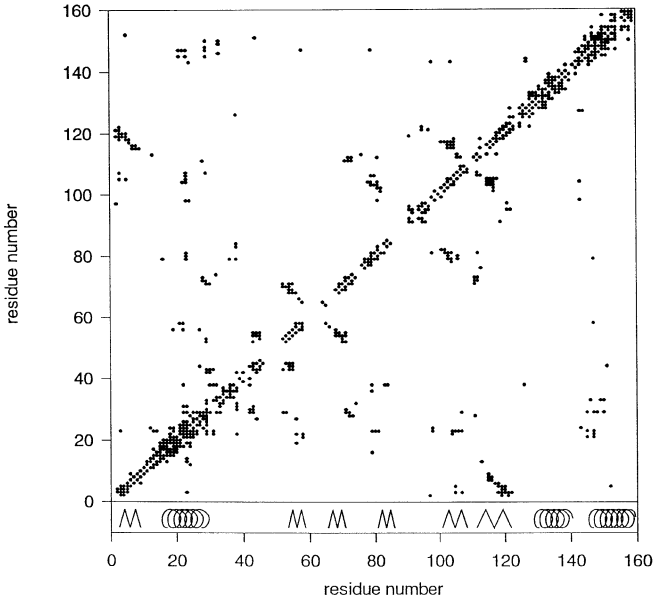

FIG. 5. **Summary of the spatial interactions between the amino acids of Bet v 1.** Each *dot* indicates that at least one NOE was observed between the corresponding residues. The *bar above* the *horizontal sequence numbers* indicates the corresponding secondary structure elements (M indicates one strand of a β-sheet).

are found at 2.30 and 2.69 ppm. In an α-helix, Gly^{26} would be situated right between the aromatic rings of Phe^{22} and Phe^{30}. This gives rise to strong shielding effects and explains the large deviation from the random coil chemical shift.

In addition to the helical regions, several strands of β-sheets can be deduced from the chemical shift analysis. These are the sections from Val^{2} to Ile^{13}, from Ile^{53} to Phe^{58}, from Asn^{100} to Thr^{107}, and from Lys^{115} to Lys^{123}. The exact start and end of each strand cannot be given with certainty. The structural elements in the middle of the sequence cannot be characterized, due to the lack of assignments. Analysis of the chemical

TABLE I

Comparison of the results of several secondary structure predictions

a), Gibrat *et al.* (40); b), Levin *et al.* (41); c), Deleage and Roux (42); d), Geourjon and Deleage (43); e), Rost and Sander (44). The results from our NMR experiments according to the chemical shift index strategy (30) (f) and the results from molecular dynamics calculations (g) are also shown. The first lines state the primary structure with the numbering of the amino acids. H indicates helical sections; E indicates strands of β-sheets; S in line b indicates bend; T in lines b and c indicates turn. A lowercase e indicates that a β-sheet cannot be safely deduced in the corresponding region due to the degeneration of chemical shifts.

```
     1       10        20        30        40        50        60        70        80
     |       |         |         |         |         |         |         |         |
     GVFNYETETTSVIPAARLFKAFILDGDNLFPKVAPQAISSVENIEGNGGPGTIKKISFPEGFPFKYVKDRVDEVDHTNFK
a    EEEE----EHHHHHHHHHHHHHHHHH-HH-HHH-HHH-EEEEEEE-------EEEE--------HH-H--HHHHHHHHH-
b    EEEE--HS--EE----HHHHHE----S----T---HHHHHHH---TT---S-E-E----TS---HHHHSHH-HH-THHHH
c    -EE-EH-HEEEEE-HHHHHHHEEHTT---H--EH-HHEEEEH-ETTT-T---EEEEETT--T-E-EEHHEEHHE----ET
d    EEE-----------HHHHHHHEEE------------HHHHHHH---------EEE--------EEEE-----HH------
e    --EEEEE-------HHHHHHHHHH----------HHHH--EEEEEE-----EEEEEEE------EEEEHHHHHH----EE
f    EEEE-EEEEEE-HHHHHHHHHHHH-HHHHH----------------------EEEEEE----EEE-EEE-EEE-----E
g    -EEEEEEE------HHHHHHHHHHHHHHHHH----------------------EEEEE--------EEEEE----------
     81      90        100       110       120       130       140       150       159
     |       |         |         |         |         |         |         |         |
     YNYSVIEGGPIGDTLEKISNEIKIVATPDGGSILKISNKYHTKGDHEVKAEQVKASKEMGETLLRAVESYLLAHSDAYN
a    E-EEEEE------EHHHHHHHHEEEE------EEEEE--------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-
b    H-EEEE-S---HHHHHHHHTHH-EE----T---EEEETT------HHHHHHHHHHHHHHHHHHHHHHHHHHHH---S--H
c    E-EEEE--------HHHE--HEEEEE--T---EEEEET--TT----HHHHHHHHH-HHH-HHHHHHEHHEEHHH--H--
d    --EEEE------HHHHHHHHHHEEEE------EEEE----------HHHHHHHHHHHHHHHHHHHHHHHHHHH-------
e    EEEEEEE----HHHHHHHH---EEEE-----EEEEEEEEE-------HHHHHHHHHHHHHHHHHHHHHHHHHH-----
f    EE-----------------EEEEEEE---EEE-EEEEEEEEEE-EHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-HHHH
g    EEEE---------------EEEEEEE----EEEEEEEEEE------HHHHHHHHHHHHH---HHHHHHHHHHHHHHH
```

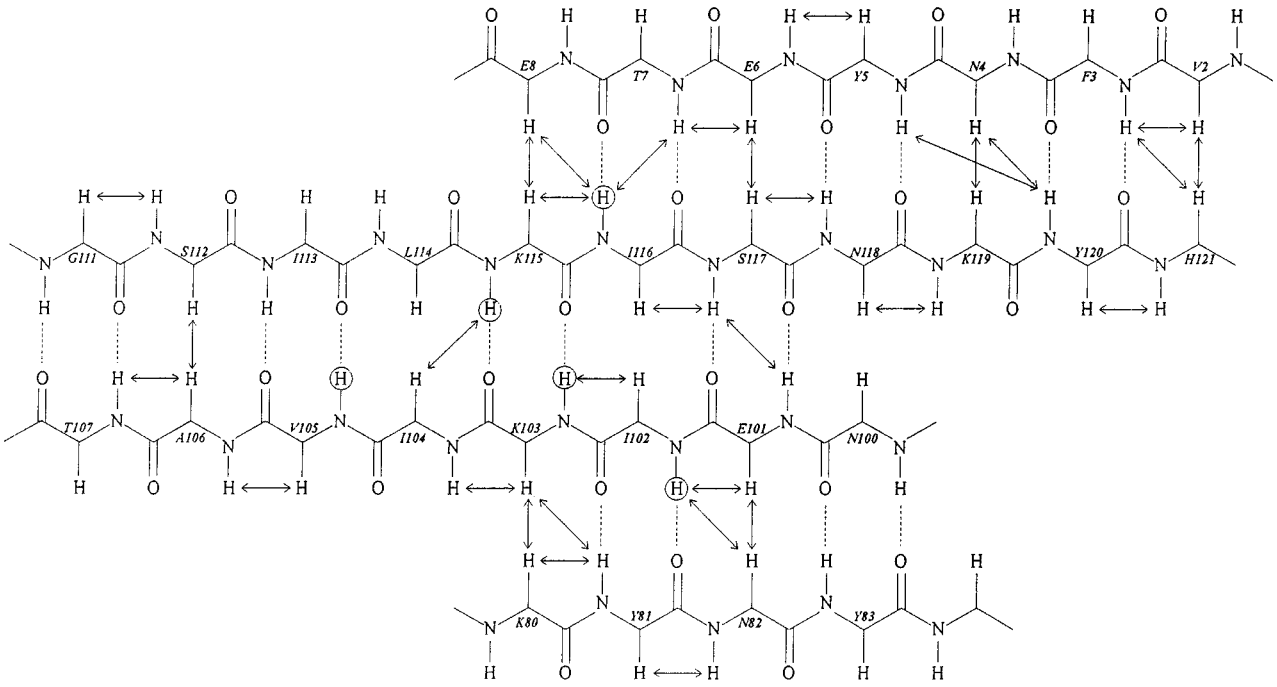

FIG. 6. **Schematic representation of the four-stranded antiparallel β-sheet.** *Arrows* indicate observed NOEs that are typical for β-sheets; *circles* indicate slow exchanging amide protons. Only NOEs that were assigned with certainty are shown. The interaction of the strands Val^{2}–Glu^{8} and Lys^{115}–His^{121} is the main reason for the globular structure of Bet v 1.

shifts only gives information about the kind of structural elements but not the relative orientation toward each other. This problem could be overcome by an analysis of all NOEs and the calculation of a crudely defined structure, although unambiguous assignment of the NOEs was in part impeded by the strong frequency degeneration. For this reason we tried to obtain unambiguous assignments of ambiguous NOEs from crudely defined three-dimensional structures in an iterative procedure.

For the structure calculation the new strategy introduced by Nilges was used (29). This method allows, in addition to the safely assigned NOEs, the use of ambiguously assigned NOEs specified directly in terms of the chemical shift assignments as described under "Experimental Procedures." All protons are taken into account that show resonances at a chemical shift where an NOE is observed. After the calculation of a set of structures all NOEs that were violated in more than 50% of the structures were discarded because they are most likely arising from protons not yet assigned. After several sets of calculated structures, 975 NOEs out of 1158 were assigned with certainty. In a contact plot (Fig. 5), several NOE patterns typical for regular secondary structure elements are easily recognized.

Bet v 1 has helical conformation from residue 15 through 31, 128 through 139, and 144 through 159. The C terminus is composed of two helices connected by a hinge at Gly^{140}. This finding is in accordance with the secondary structure predic-

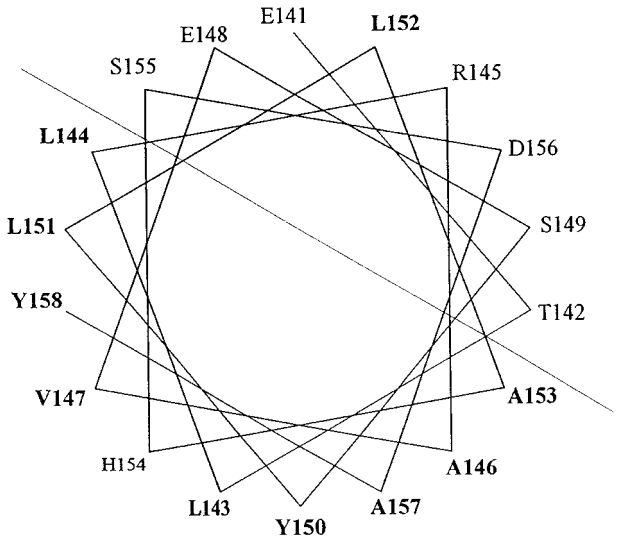

FIG. 7. **Helical wheel presentation of the C-terminal helix of Bet v 1 showing the amphipathic character.** The *diagonal line* separates the hydrophobic from the hydrophilic side. Hydrophobic amino acids are in *boldface type*.

tions as well as with the results from the secondary structure determination by the chemical shift index strategy (Table I). Detailed information about the helices could not be obtained due to the frequency degeneration of the signals of the aliphatic side-chain protons.

Bet v 1 contains six strands of antiparallel β-sheets. A four-stranded β-sheet is composed of sections Val^2–Glu^8, Lys^{80}–Tyr^{83}, Asn^{100}–Thr^{107}, and Gly^{111}–His^{121}. The arrangement of the four strands is shown in Fig. 6. *Arrows* indicate observed and assigned backbone NOEs typical for a β-sheet, and *circles* indicate slowly exchanging amide protons. This four-stranded β-sheet dominates the global fold of Bet v 1 and is therefore responsible for the globular structure of the protein. The two strands Val^2–Glu^8 and Gly^{111}–His^{121} fix both the N terminus and the beginning of the helix (Val^{128}–Met^{139}).

Another double-stranded β-sheet is composed of sections Ile^{53}–Ser^{57} and Tyr^{66}–Arg^{70}. The orientation of the double-stranded β-sheet and the existence of a third strand around Asn^{43} could not be unambiguously determined due to the incomplete assignment of the loop regions connecting the strands. New information from double-labeled Bet v 1 will lead to clarification of this problem as well as to the determination of the orientation of the helices and the β-sheets toward each other. The results from full-length Bet v 1 suggest that both C-terminal T cell epitopes (145–156 and 147–158) are situated in an amphipathic α-helix (Fig. 7). In order to obtain further information about structure and stability of these two C-terminal helices, a synthetic Bet v 1-(125–154) peptide was investigated. The N terminus of the peptide was chosen to contain the complete first C-terminal helix, as deduced from the full-length Bet v 1. The exact length of the peptide (Asp^{125}–His^{154}) was chosen for reasons of good solubility, in spite of the fact that both T cell epitopes stretch further toward the C terminus.

The isolated peptide shows only a weak tendency to form helical secondary structure in aqueous solution, as monitored by CD spectroscopy (8). The content of secondary structure was clearly enhanced upon the addition of 60% trifluoroethanol. Taking as a working hypothesis that the addition of TFE compensates for the decreased secondary structure stability due to the loss of interactions with other parts of the protein (37), we decided to determine the structure in 60% TFE solution.

For Bet v 1-(125–154) all amino acid residues except Glu^{17}(141) (numbering scheme of Bet v 1 given in parentheses) could be assigned with the usual strategies (12). Most resonances of the C$^\alpha$ protons are shifted toward higher field as compared with random coil values, indicating helical conformation. 167 interresidual and 306 intraresidual NOEs were used for structure calculations. None of these NOEs was violated in the resulting 20 structures (Table II). Fig. 8 gives an overview of the observed sequential and medium range NOEs. For residues Glu^3(127)–Gln^8(132), $d_{\alpha N}$(i, i + 3) and $d_{\alpha\beta}$(i, i + 3) NOEs

TABLE II
Energy contributions to the structure and deviations from standard geometry

E_{VDW}, van der Waals energy; E_{NOE}, effective NOE energy term. The energy values reported here result from the target function used for the molecular dynamics calculations as described in the text. All values are mean values over 20 refined structures. Values in parentheses indicate the standard deviations from this mean.

	Average energies
	kJ/mol
E_{tot}	−407 ± 132
E_{NOE}	45 ± 20
E_{bond}	47 ± 8
E_{angle}	97 ± 17
E_{vdw}	−485 ± 138
	RMSDs from idealized geometry
NOE violations	0
NOE (nm)	0.007 ± 0.002
Bond length (nm)	0.001 ± 0.000
Bond angles (deg)	0.855 ± 0.141
Impropers (deg)	0.754 ± 0.130

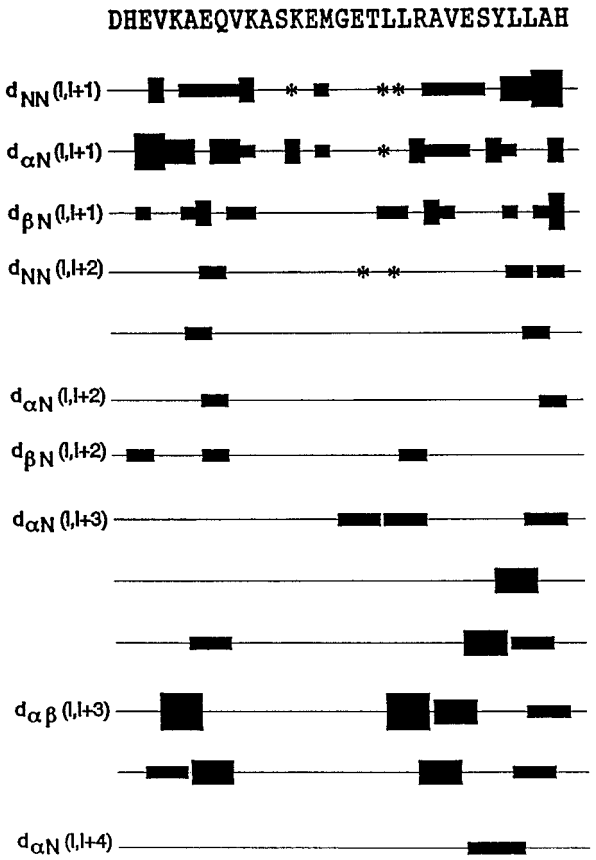

FIG. 8. **Pattern of NOESY cross-peaks as derived from a 200-ms NOESY spectrum of Bet v 1-(125–154).** The *thickness* of the *bars* qualitatively indicates the relative strength of the cross-peaks. An *asterisk* indicates that the NOE could not be observed because of frequency degeneration. The *first line* states the amino acid sequence.

could be detected, and for residues $Gly^{16}(140)$–$His^{30}(154)$, $d_{\alpha N}(i, i + 3)$, $d_{\alpha\beta}(i, i + 3)$, and $d_{\alpha N}(i, i + 4)$ NOEs were observed. Consequently, Bet v 1-(125–154) shows α-helical conformation in sections $Glu^{3}(127)$–$Val^{9}(133)$ and $Gly^{16}(140)$-$His^{30}(154)$. The section from $Lys^{10}(134)$ to $Gly^{16}(140)$ is flexible and acts as a hinge. The average RMSD for the helical sections is very low (0.04 nm) compared with the RMSD in the hinge region of about 0.1 nm (Fig. 9). Fig. 10 shows an overlay of 10 structures for each of the two helical segments.

The structure of Bet v 1-(125–154) is in very good accord with the results from full-length Bet v 1. This finding underlines the result that TFE is able to stabilize a native-like structure (38). For example, helix stop signals remain intact (39). Both segments that were found to be helical in Bet v 1 clearly have α-helical conformation in the peptide, underlining the amphipathic character of the T cell epitopes (Fig. 7). The relative orientation of the helices toward the rest of the protein can be estimated by inspection of the contact plot of full-length Bet v 1 given in Fig. 5.

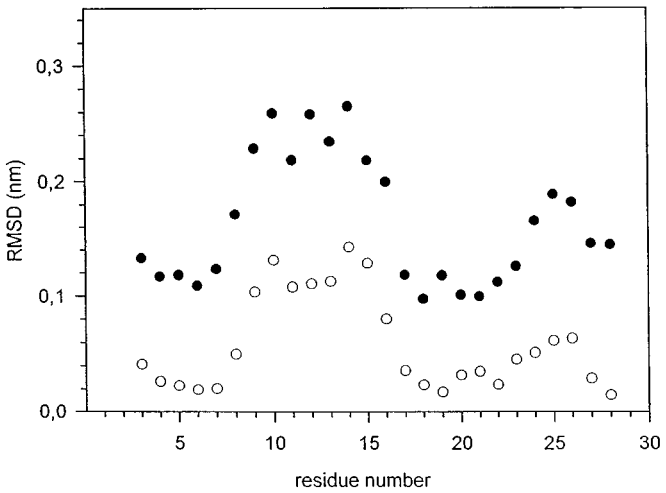

FIG. 9. **Diagram of the RMSD values per residue for the 20 structures obtained after molecular dynamics calculations.** The RMSD values for the backbone atoms and all heavy atoms are represented by *hollow* and *filled circles*, respectively.

The enlargement of the contact plot for residues 80–140 of Bet v 1 shows striking similarity with the corresponding contact plot for the ragweed allergen Amb t V (5). The structure of Amb t V is mainly composed of a triple-stranded antiparallel β-sheet connected to a C-terminal helix via a turn. The same arrangement is also found in Bet v 1, only the β-sheet is supplemented by a fourth antiparallel strand from the N terminus, and the connecting loop regions are longer than for Amb t V. Even the relative orientation of the secondary structure elements is similar. Two long range NOEs between the last strand of the β-sheet and the end of the helix are observed. Thus, in spite of the totally different amino acid sequence both allergens show identical folds. In Amb t V, the common structural motif is mainly stabilized by four disulfide bridges, whereas in Bet v 1 the stabilization is effected by the globular spatial structure and the resultant originating interactions between all parts of the protein.

The section of Bet v 1 that has a structural equivalent in Amb t V is between residues 80 and 140. There are so far no indications that antibody binding is specific for a certain type of tertiary structure. Therefore, this structural similarity does not necessarily hint at a function of this specific three-dimensional structure in IgE binding, but a function common to Amb t V and Bet v 1 cannot be ruled out for this segment. The T cell epitope found between Ile^{113} and Gly^{124} may possibly serve as a B cell epitope as well. The T cell epitopes in the C-terminal helix (positions 145–156 and 147–158) have no equivalent in Amb t V, because they are situated in the second C-terminal helix and thus not close to the β-sheet-turn-helix motif found in Amb t V. Therefore, we hypothesize that the C-terminal helix yields the sequences for T cell epitopes, but the B cell epitope is situated near the end of the β-sheet and the beginning of the penultimate helix. Corroboration of this hypothesis will be brought from the determination of the exact structure of Bet v 1 with ^{1}H–^{15}N–^{13}C spectroscopy and from the determination of

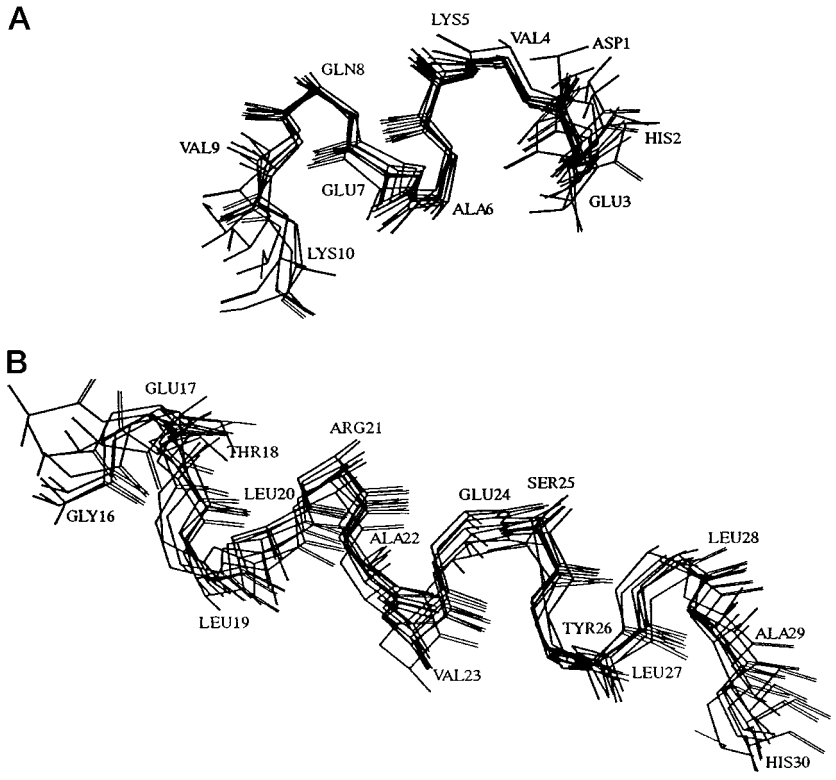

FIG. 10. **Overlay of 10 structures for each of the helical segments of the Bet v 1-(125–154) peptide.** The overlays from Glu^{3} to Lys^{10} and from Gly^{16} to His^{30} are shown in *A* and *B*, respectively. The structures showed the lowest energy after the molecular dynamics calculations and the best agreement with the experimental data.

structures of other allergens.

Another question that is still unclear is the physiological function of Bet v 1 in birch pollen. In order to identify possible structure-function relationships, a search for homologies on the level of secondary structure was performed using the TOPITS (34) algorithm. The data base search revealed that the motifs of secondary structure found in Bet v 1 are highly similar to those found in guanosine nucleotide-binding proteins, including elongation factor Tu, Ras p21, purine nucleotide phosphorylase, and guanylate kinase. In addition, similarities on the level of secondary structure were detected between Bet v 1 and the endonucleases *Eco*RI and *Eco*RV.

Acknowledgments—Franz Herrmann assisted with the NDEE program. The X-PLOR simulated annealing protocol was provided by Michael Nilges (EMBL, Heidelberg).

REFERENCES

1. Breiteneder, H., Pettenburger, K., Bito, A., Valenta, R., Kraft, D., Rumpold, H., Scheiner, O., and Breitenbach, M. (1989) *EMBO J.* **8,** 1935–1938
2. Breiteneder, H., Ferreira, F., Hoffmann-Sommergruber, K., Ebner C., Breitenbach, M., Rumpold, H., Kraft, D., and Scheiner, O. (1993) *Eur. J. Biochem.* **212,** 355–362
3. Larsen, J. N., Störman, P., and Ipsen, H. (1992) *Mol. Immunol.* **29,** 703–711
4. Breiteneder, H., Ferreira, F., Reikertsorfer, A., Duchene, M., Valenta, R., Hoffman-Sonnengruber, K., Ebner, C., Breitenbach, M., Kraft, D., and Scheiner, O. (1992) *J. Allergy Clin. Immunol.* **90,** 909–917
5. Metzler, W. J., Valentine, K., Roebber, M., Friedrichs, M. S., Marsh, D. G., and Mueller, L. (1992) *Biochemistry* **31,** 5117–5127
6. Margalit, H., Spouge, J. L., Cornette, J. L., De Lisi, C., and Berzofsky, J. A. (1987) *J. Immunol.* **138,** 2213–2229
7. Ebner, C., Szepfalusi, Z., Ferreira, F., Jilek, A., Valenta, R., Parronchi, P., Maggi, E., Romangnani, S., Scheiner, O., and Kraft, D. (1993) *J. Immunol.* **150,** 1047–1054
8. Lindemann, A. (1995) *NMR-Spektroskopie zur Strukturaufklärung des viralen Proteins DsbA und des Birkenpollenallergens Bet v 1,* Ph.D. thesis, University of Bayreuth, Germany
9. Ferreira, F., Hoffman-Sommergruber, K., Breiteneder, H., Pettenburger, K., Ebner, C., Sommergruber W., Steiner, R., Bohle, B., Sperr, W. R., Valent, P., Kungl, A. J., Breitenbach, M., Kraft, D., and Scheiner, O. (1993) *J. Biol. Chem.* **268,** 19574–19580
10. Frank, R., and Gausepohl, H. (1988) in *Modern Methods in Protein Chemistry* (Tschesche, H., ed) Vol. 3, pp. 41–60, de Gruyter, Berlin
11. Ernst, R. R. (1992) *Angew. Chem. Int. Ed. Engl.* **104,** 817–852
12. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Liss, Inc., New York
13. Müller, L. (1979) *J. Am. Chem. Soc.* **101,** 4481–4484
14. Bax, A., Griffey, R. H., and Hawkins, B. L. (1983) *J. Magn. Reson.* **55,** 301–312
15. Gronenborn, A. M., Bax, A., Wingfield, P. T., and Clore, M. G. (1989) *FEBS Lett.* **243,** 93–98
16. Gronenborn, A. M., Wingfield, P. T., and Clore, G. M. (1989) *Biochemistry* **28,** 5081–5089
17. Marion, D., Discroll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry* **28,** 6150–6156
18. Marion, D., Kay, L. E., Sparks, S. W. Torchia, D. A., and Bax, A. (1989b) *J. Am. Chem. Soc.* **111,** 1515–1517
19. Fesik, S. W., Gampe, R. T., Zuiderweg, E. R. P., Kohlbrenner, W. E., and Weigl, D. (1989) *Biochem. Biophys. Res. Commun.* **159,** 842–847
20. Zuiderweg, E. R. P., and Fesik, S. W. (1989) *Biochemistry* **28,** 2387–2391
21. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113,** 967–974
22. Güntert, P., and Wüthrich, K. (1992) *J. Magn. Reson.* **96,** 403–407
23. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR* **5,** 67–81
24. Clore, G. M., Gronenborn, A. M., Nilges, M., and Ryan, C. A. (1987) *Biochemistry* **26,** 8012–8023
25. Wüthrich, K., Billeter, M., and Braun, W. (1983) *J. Mol. Biol.* **169,** 949–961
26. Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., and Wüthrich, K. (1987) *J. Mol. Biol.* **196,** 611–639
27. Qi, P. X., Di Stefano, D. L., and Wand, A. J. (1994) *Biochemistry* **33,** 6408–6417
28. Brünger, A. (1993) *X-PLOR*, 3.1 Manual, Yale University Press, New Haven, CT
29. Nilges, M. (1995) *J. Mol. Biol.* **245,** 645–660
30. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry* **31,** 1647–1651
31. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637
32. Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Petterson, G., Knowles, J., and Gronenborn, A. M. (1989) *Biochemistry* **28,** 7241–7257
33. Rost, B., Sander, C., and Schneider, R. (1994) *Comput. Appl. Biosci.* **10,** 53–60
34. Rost, B. (1995) in *The Third International Conference on Intelligent Systems for Molecular Biology (ISMB), Cambridge, UK* (Rawlings, C., Clark, D., Altman, R., Hunter, L., Lengauer, T., and Wodak, S., eds) pp. 314–321, AAAI Press, Menlo Park, CA
35. Schmidt, J. M., Thüring, H., Werner, A., Rüterjans, H., Quaas, R., and Hahn, U. (1991) *Eur. J. Biochem.* **197,** 643–653
36. Fesik, S. W., and Zuiderweg, E. R. P. (1988) *J. Magn. Reson.* **78,** 588–593
37. Zhou, N. E., Zhu, B. Y., Sykes, B. D., and Hodges, R. S. (1992) *J. Am. Chem. Soc.* **114,** 4320–4326
38. Sönnichsen, F. D., van Eyk, J. E., Hodges, R. S., and Sykes, B. D. (1992) *Biochemistry* **31,** 8790–8798
39. Nelson, J. W., and Kallenbach, N. R. (1989) *Biochemistry* **28,** 5256–5261
40. Gibrat, J. F., Garnier, J., and Robson, B. (1987) *J. Mol. Biol.* **198,** 425–443
41. Levin, J. M., Robson, B., and Garnier, J. (1986) *FEBS Lett.* **205**, 303–308
42. Deleage, G. and Roux, B. (1987) *Protein Eng.* **1,** 289–294
43. Geourjon, C., and Deleage, G. (1994) *Protein Eng.* **7,** 157–164
44. Rost, B., and Sander, C. (1994) *Proteins* **19,** 55–72