J. Biochem. 2010;147(1):127–133 doi:10.1093/jb/mvp150

JB THE JOURNAL OF BIOCHEMISTRY

Characteristic amino acid distribution around segments unique to allergens

Received July 1, 2009; accepted September 8, 2009; published online October 8, 2009

Naoyuki Asakawa[1,*], Noriyuki Sakiyama[1,2,†], Reiko Teshima[3] and Shigeki Mitaku[1]

[1]Department of Applied Physics, School of Engineering, Nagoya University, Chiku-sa-ku Frocho, Nagoya; [2]Venture Business Laboratory, Nagoya University, Chiku-sa-ku Frocho, Nagoya, Aichi, 464-8603; and [3]Novel Foods and Immunochemistry Division, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan

*Present address: Computational Science Department, Science & Technology Systems Division, Ryoka Systems, Inc., Shinkawa, Chuo-ku, Tokyo, 104-0033, Japan

†Noriyuki Sakiyama, Department of Applied Physics, School of Engineering, Nagoya University, Chiku-sa-ku Frocho, Nagoya, Aichi, 464 86073, Japan, Tel: +81-52-789-4465, Fax: +81-52-789-3706, E-mail: sakiyama@bp.nuap.nagoya-u.ac.jp

Epitopes are located at the surface of allergens with which antibodies specifically bind. On the assumption that fragments unique to allergens have common, characteristic amino acid sequences, we compared the amino acid sequences of allergens with those of non-allergens. Segments around fragments unique to allergens showed wavelet-like distributions for several amino acids. Charged residues, alanine and glycine had positive peaks at the centre of the unique segments with small valleys on both sides, while aromatic residues, proline and cysteine showed the inverse distribution. Furthermore, the wavelet-like distribution of amino acids could be represented by a universal distribution function together with an index characterizing the intensity of the wavelet. Using the universal distribution function and the novel index of amino acids, we developed a simple method for extracting segments and fragments that are unique to allergens. The significance of the universal distribution function and the novel index is also discussed, by comparing the plot of the allergen-unique fragments index and dynamic fluctuation in the three dimensional structure of birch pollen allergen as both a single molecule and a complex with the corresponding antibody.

Keywords: allergen/unique sequence fragment/protein–protein interaction/bioinformatics.

Abbreviations: AUFs, allergen-unique fragments; ALG dataset, allergen dataset; NEG dataset, Negative datasets of non-allergen proteins; EPITOPE dataset, IgE epitope-known dataset; PDB, Protein databank; ALG fragments, fragments recurring in only allergen proteins; NEG fragments, fragments recurring in only non-allergen proteins; BOTH fragments, fragments recurring in both types of protein.

Most proteins are composed of compact core regions and dynamically fluctuating segments. The fluctuating parts in a protein or a protein complex can be identified by the B-factor of atoms on X-ray crystallography. Most fluctuating segments are located at the surface of proteins and play an important role in protein–protein interactions (*1–4*). Therefore, the relationship between the degree of fluctuation in the three-dimensional structures and the characteristics of amino acid sequences is an interesting and important problem to be solved. The characteristics of amino acid segments of completely disordered regions in proteins have recently been studied by several groups (*5–12*), but the relationship between amino acid sequences and the degree of fluctuation of segments in ordered structures has not been studied in detail (*13*).

Environmental effects on the structural fluctuation can lead to difficulties in studying the relationship between amino acid sequences and dynamic fluctuation. For example, large fluctuations in a segment within a single molecule are suppressed by binding with another molecule (*2*, *3*). Similarly, an intrinsically flexible segment in a domain becomes rigid upon binding with another domain in the same protein. Proteins thus contain numerous segments whose fluctuations are intrinsically large, but are suppressed in the native state or in protein complexes. Therefore, it is very difficult to develop a computational method for predicting the degree of dynamic fluctuation in amino acid segments.

Antigen–antibody interactions provide an interesting problem with regard to the dynamic structures of amino acid segments. There are several epitopes on the surface of antigens that are flexible and bind with antibodies (*3*). Allergen proteins are a category of antigens that interact with immunoglobulin E (IgE). It is thought that fragments unique to allergens, and that are not found in non-allergen proteins, are located at the surface of these proteins (*14*). Fragments unique to allergens that have potential flexibility for binding with IgE are thus thought to exist.

In the present study, we analysed the amino acid sequences of 663 allergens and 539 non-allergens and obtained 8,696 fragments unique to allergens. The distribution of amino acids around the allergen-unique fragments (AUFs) was then calculated, revealing the wavelet-like distribution of several amino acids. In order to elucidate the meaning of the wavelet-like amino acid distribution, we developed a novel index of amino acids (AUF index) that represents the intensity of wavelet-like distribution. When the plots of the AUF index were studied for the birch pollen allergen, the peaks correlated well with the portions of large structural fluctuations. The possibility of developing a software system to predict amino acid segments having intrinsically

enhanced fluctuation and epitopes of allergens is also discussed.

Materials and methods

Datasets of allergens and non-allergens

The allergen dataset (ALG dataset) was prepared from the Allermatch (*15*), and 663 allergen proteins were obtained, omitting redundant sequences with point mutations. Negative datasets of non-allergen proteins (NEG dataset) consisting of 539 proteins were obtained from the literature (*14*), omitting identical sequences and fragment entries. The IgE epitope-known dataset (EPITOPE dataset) was derived from SDAP (*16*), the CSL Allergen Database (*17*) and ProtAll (*18*). Datasets are summarized in Table 1.

Data on three-dimensional structures of birch pollen allergen in both the single-molecule state (PDB: 1BV1) (*19*) and as a complex with the antibody at the IgE epitope (PDB: 1FSK) (*20*, *21*) were obtained from the Protein databank.

Dataset of AUFs

We prepared AUFs in three steps. First, we searched for three to eight amino acid fragments recurring in each protein from the allergen and non-allergen protein datasets. Second, fragments were classified into 4 categories: fragments recurring in only allergen proteins (ALG fragments); fragments recurring in only non-allergen proteins (NEG fragments); fragments recurring in both types of protein (BOTH fragments); and fragments occurring only once in proteins. The number of ALG fragments, NEG fragments and BOTH fragments was respectively 4297, 2021 and 5155. In the third step, taking into account the fact that many of epitopes of allergen consist of pairs of fragments, we prepared fragment pairs that were unique to allergen proteins, combining ALG fragments and BOTH fragments. Although the length of fragment pairs can be very large, we used the segments covered by the fragment pairs, whose length was 3–15 residues. We defined those segments as AUFs. Figure 1 shows the length distribution of the AUFs which indicated that the length of AUFs monotonously decreased, and the segments shorter than 15 residues are enough for further study of the characteristics of the segments. The number of AUFs was 8696. When 57 epitope segments in 16 allergen proteins were analysed by alignment with AUFs, 53 epitopes overlapped with AUFs, covering 93% of epitopes.

Table 1. Sequence datasets for allergens, non-allergens and epitopes.

Datasets	No. of sequences	Databases
ALG	663	Allermatch (*15*)
NEG	539	Zorzet *et al.* (*14*)
Epitope	57	SDAP (*16*), Allergen Database (*17*), ProtAll (*18*)

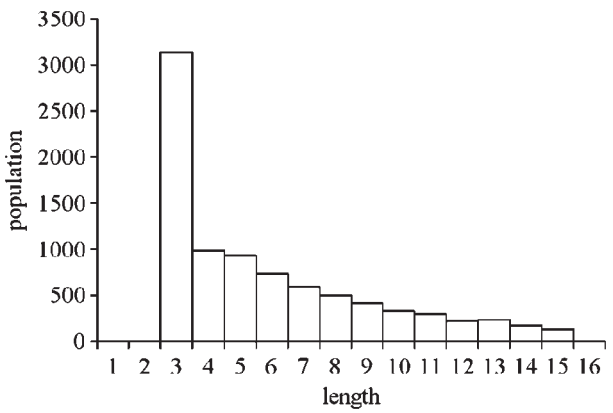

Fig. 1 The distribution of length of fragments in AUF.

Frequencies of appearance of amino acids around the centre of AUFs

We prepared AUFs for extracting fluctuating fragments at the surface of allergens. AUFs are not necessarily epitopes, but must contain epitope-like fragments, as epitopes have the following characteristics (*1*, *3*, *22*): (i) the binding portion, epitope, of an allergen consists of several fragments that are exposed to the surface of the protein, and at least one of the fragments that constitute the binding portion is unique to an allergen, guaranteeing the uniqueness of the binding portion; and (ii) there are no common sequence motifs throughout all epitopes, and the binding portion has a dynamic state for binding with the corresponding IgE. The dataset of AUFs recurring in amino acid sequences of allergens would therefore be useful for developing software tools to predict allergens, as well as for elucidating the mechanisms of molecular recognition by immunoglobulins.

In order to elucidate the common characteristic distributions of amino acids around the centre of AUFs, we calculated the frequencies of appearance of amino acids at each position of 29 amino acids in which AUFs are located at the central positions. When a fragment had odd number of amino acids, the count was raised by one to the distribution of each amino acid at the corresponding position. When a fragment had even number of amino acids, we considered a virtual sequence with odd number of residues in which the two amino acids neighbouring to the virtual residue equally contribute by 0.5 to the virtual position. The amino acid distributions around the centre of fragments in ALG, NEG and BOTH were obtained from the same way as AUFs.

AUF plot for predicting AUF-like segments

Using the universal distribution function in Fig. 3B and the AUF index in Table 2, we developed an AUF plot expressed by the following equation,

$$S_{\mathrm{AUF}}(i) = \sum_{j=-10}^{10} A(i+j) \cdot f(j) \tag{1}$$

Here, $S_{\mathrm{AUF}}(i)$ represents the score of a segment of 21 residues around the i-th position of an amino acid sequence. Function $f(j)$ represents the universal distribution function for AUFs in Fig. 3B, in which j indicates the interval from the centre of the wavelet. Coefficient $A(i + j)$ is the AUF index value, corresponding to the amino acid at the $(i + j)$ position. A peak on the AUF plot represents the degree to which a segment around the peaks has the characteristics of the AUF.

Average hydrophobicity around peaks of AUF plots

Hydrophobic segments generally form the core of proteins, while hydrophilic segments are located at the periphery. To evaluate the hydrophobicity around the peaks of AFU plots, we calculated the double moving average $\langle\langle H(i)\rangle\rangle$ of the Kyte–Doolittle hydropathy index (*23*) $H(k)$. We adopted the double average in order to smooth very notched hydropathy index sequences.

$$\langle\langle H(i)\rangle\rangle = \sum_{j=i-3}^{i+3} \frac{\left(\sum_{k=j-3}^{j+3} H(k)/7\right)}{7} \tag{2}$$

Table 2. Amino acid index characterizing distribution of allergen-unique segments.

Positive weight		Negative weight		No weight	
D	170.0	M	−201.5	I	0.0
E	224.5	W	−161.0	V	0.0
A	238.5	F	−159.0	L	0.0
G	253.5	C	−158.5	T	0.0
K	285.0	Y	−152.0	S	0.0
		H	−141.0	N	0.0
		P	−90.0	Q	0.0
				R	0.0

Results

Characteristic distribution function of amino acids around AUFs

The distribution of amino acids around AUFs was classified into three categories: wavelet-like distribution with a peak at the centre of AUFs; inverse distribution with a valley at the centre; and random distribution. Figure 2A–C shows the distributions of the three categories of amino acid. Glycine, alanine, aspartic acid, glutamic acid and lysine exhibited distributions with a peak at the centre (Fig. 2A), while phenylalanine, histidine, tyrosine, tryptophan, cysteine, methionine and proline belonged to the second category having a valley at the centre (Fig. 2B). Random distribution around AUFs was observed for leucine, isoleucine, valine, serine, threonine, arginine, asparagine and glutamine (Fig. 2C).

The amino acid distributions around single fragments of ALG, NEG and BOTH were analysed and classified in three groups in the same way as AUFs. The distribution with a peak and a valley at the centre of BOTH fragments showed the same tendency of Figure 2A–C. The distributions of amino acids for AUFs were similar to those for ALG and BOTH data set but showed large different from those for NEG. The details of the difference in the wavelet-like distribution among the data set for ALG, NEG, BOTH and AUFs provide interesting problem from the aspect of the epitope prediction, but we used the distributions of amino acids around AUFs for simplifying the analysis of many amino acid sequences. The detailed analysis of the three types of data set (ALG, NEG, BOTH) will be described elsewhere.

The distribution of each amino acid could be expressed by the wavelet-like distribution with the offset of random distribution, and the shapes of the wavelet-like part of the distributions for amino acids in the same categories were very similar. Figure 3A shows the average normalized distributions of the wavelet-like parts for each category of amino acid, having a peak or a valley at the centre. When one of the normalized distributions was inversed vertically, the two graphs overlapped. Glycine, alanine, aspartic acid, glutamic acid and lysine were used for calculating the normalized distributions of a peak at the centre, and phenylalanine, tyrosine, tryptophan, cysteine and methionine were used for the normalized distributions of a valley. Because the valley for proline was not significant, proline was not used for the calculation of the normalization. We then obtained the universal distribution of amino acids around AUFs by averaging the distribution of the first category and the inversed distribution of the second (Fig. 3B). Table 2 shows the intensities of the wavelet-like distribution for amino acids that were used in this study as the AUF index. The amino acids in the first category showed positive values for the AUF index, while amino acids in the second category had negative values. The value of the index was zero for the amino acids in the third category.

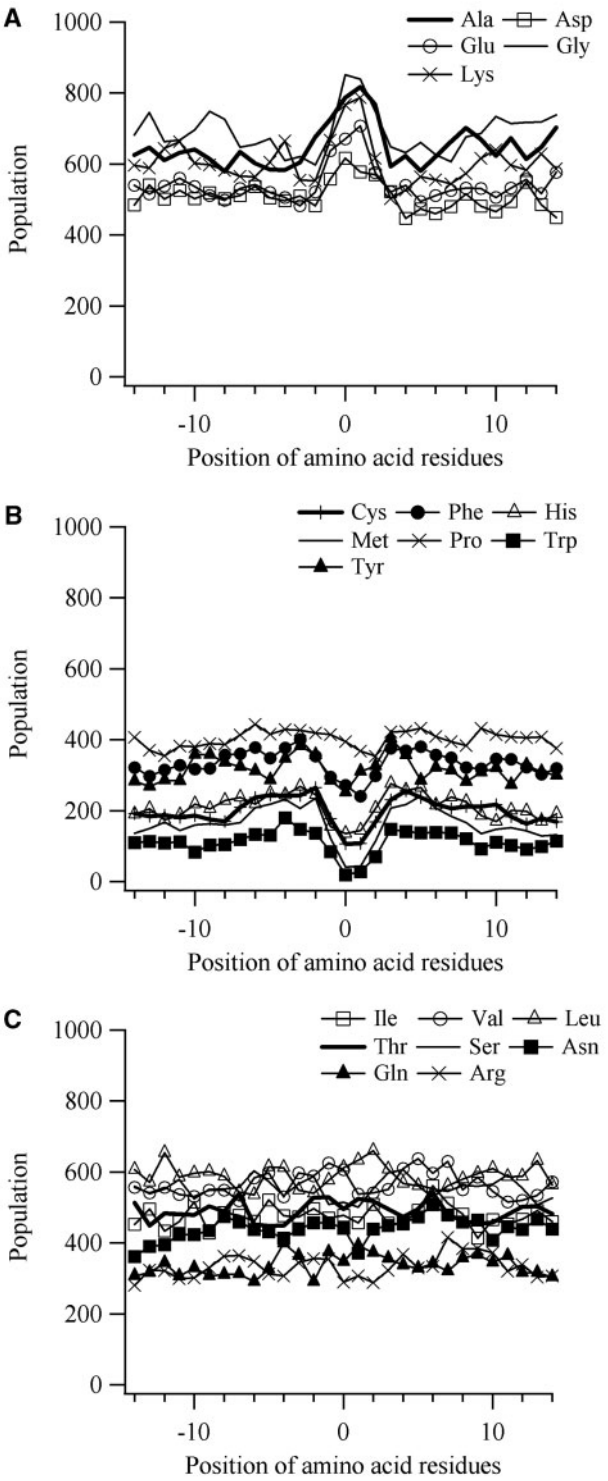

Fig. 2 Three types of amino acid distribution around allergen-unique sequences: peak at the centre (A), valley at the centre (B) and random distribution (C). Glycine, alanine, aspartic acid, glutamic acid and lysine belong to the first category, and phenylalanine, histidine, tyrosine tryptophan cysteine, methionine and proline are the members of the second category. Other amino acids showed random distribution (third category).

Correlation between peaks of AUF plot, and positions of secondary structures and epitopes

In order to better understand the AUF plot, we analysed the sequence of an allergen; the birch pollen allergen. We studied this allergen because it is the only

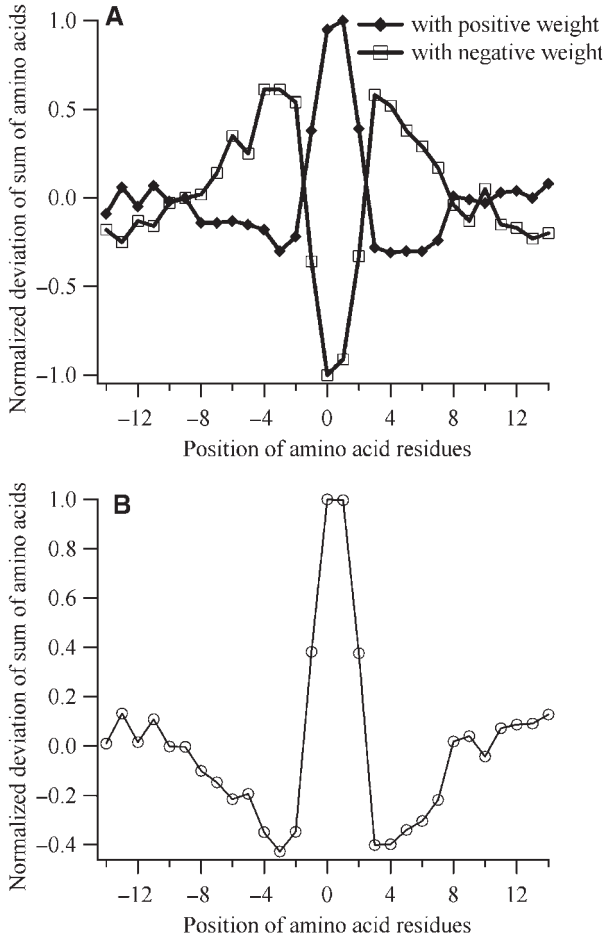

Fig. 3 Wavelet-like average distribution of amino acids for the first category have a peak at the centre with two small valleys on both sides, while wavelet-like distribution for the secondary category has the inverse shape (A). Averaging the distribution for the first category and the inverse distribution of the secondary category, the universal curve is obtained for all wavelet-like distribution of amino acids around allergen-unique segments (B).

one whose molecular structures are known both in the single-molecule state and in the complex with the corresponding IgE. An AUF plot for the birch pollen allergen is shown in Fig. 4A (Graph I). The secondary structure of the birch pollen allergen (PDB: 1BV1) is also shown in Fig. 4A (Graph II). It appears that the peaks of the AUF plot are positioned at the loops or the end of the secondary structure. It is clear that the peaks of the AUF plot correspond to the loops or the end regions of the secondary structures, thus suggesting that the peaks are dynamically fluctuating segments at the surface of the protein. Previous work on the secondary structure breakers showed that proline, glycine and amphiphilic residues, such as lysine and glutamic acid are signals for the loops or ends of secondary structures (*24*). These residues are, in fact, the main factors for the peaks in the AUF plots.

Graph III in Fig. 4A shows the hydrophobicity plot of the birch pollen allergen. We then produced three graphs (IV, V and VI) of the peaks of the AUF plot for visualizing the positions of peaks in the hydrophobic regions $\langle\langle H(i)\rangle\rangle > 0.0$ (Graph IV), the moderately hydrophilic region $-1.5 < \langle\langle H(i)\rangle\rangle < 0.0$ (Graph V) and the very hydrophilic region $\langle\langle H(i)\rangle\rangle < -1.5$ (Graph VI). Peaks in region IV are indicated in blue and green; the blue peaks correspond to higher hydrophobic regions ($\langle\langle H(i)\rangle\rangle > 0.5$), whereas the green peaks correspond to regions with lower hydrophobicity ($0.0 < \langle\langle H(i)\rangle\rangle < 0.5$). Only one peak was found in region VI, which in indicated in red. All other peaks were in region V, indicating that peaks are mainly present in moderately hydrophilic regions. Graph VII shows the epitope regions (*21*). Comparison of epitope regions and Graphs IV, V and VI shows that the epitopes in this allergen correspond to the peaks in the AUF plot with moderate hydrophilicity.

The three-dimensional structures of the birch pollen allergen from two directions are shown in Fig. 4B. The positions of the AUF peaks are marked by coloured letters from a to v. Four colours (blue, green, red and orange) indicate the hydrophobic regions used in Fig. 4A. From the physicochemical viewpoint, it is unsurprising that the hydrophobic AUF peaks (blue and green) are clustered, forming a hydrophobic core in this protein, whereas the moderately hydrophilic AUF peaks are located at the peripheral regions. AUF peaks corresponding to epitopes are indicated by f, g, i, j, l and n, which are clustered on one side of the protein.

Correlation between peaks of AUF plot and normalized B-factor

The three-dimensional structures of proteins in both the single-molecule state and the antibody complex are now available for the birch pollen allergen (*19–21*). Therefore, the changes in structural fluctuation upon binding with IgE may be visualized by subtracting the B-factors of the compared positions of the AUF peaks. Figure 5A shows the plots of B-factors for the structures of birch pollen allergen in the single-molecule (Graph VIII) and antibody complex (Graph IX) states, as normalized by Smith's method (*25*). The dotted lines in Fig. 5A show the positions of the AUF peaks. As expected from the nature of the protein–protein binding, the fluctuation of fragments at the epitopes (Graphs VIII and IX) is greatly reduced by complex formation between the allergen and IgE. Graph X shows the difference in B-factors between the two states. The results indicate that the segment between the 25-th and 80-th residues, indicated with downward arrows, shows great reductions in structural fluctuation, corresponding to the main part of the epitopes. On the other hand, fluctuation of the segments around the 110-th, 130-th and 150-th residues is significantly enhanced by protein–protein binding. Figure 5B shows the structure of birch pollen allergen, in which the difference in B-factors between the single-molecule and protein complex states is plotted using pseudo-colours. Fluctuations in the segments near the epitopes in the sequence are reduced significantly, while the segments far from the epitopes show enhanced fluctuation by binding with IgE. The fact that the fluctuation is suppressed by protein–protein interactions is reasonable from a physical standpoint, and the correlation between the positions of AUF

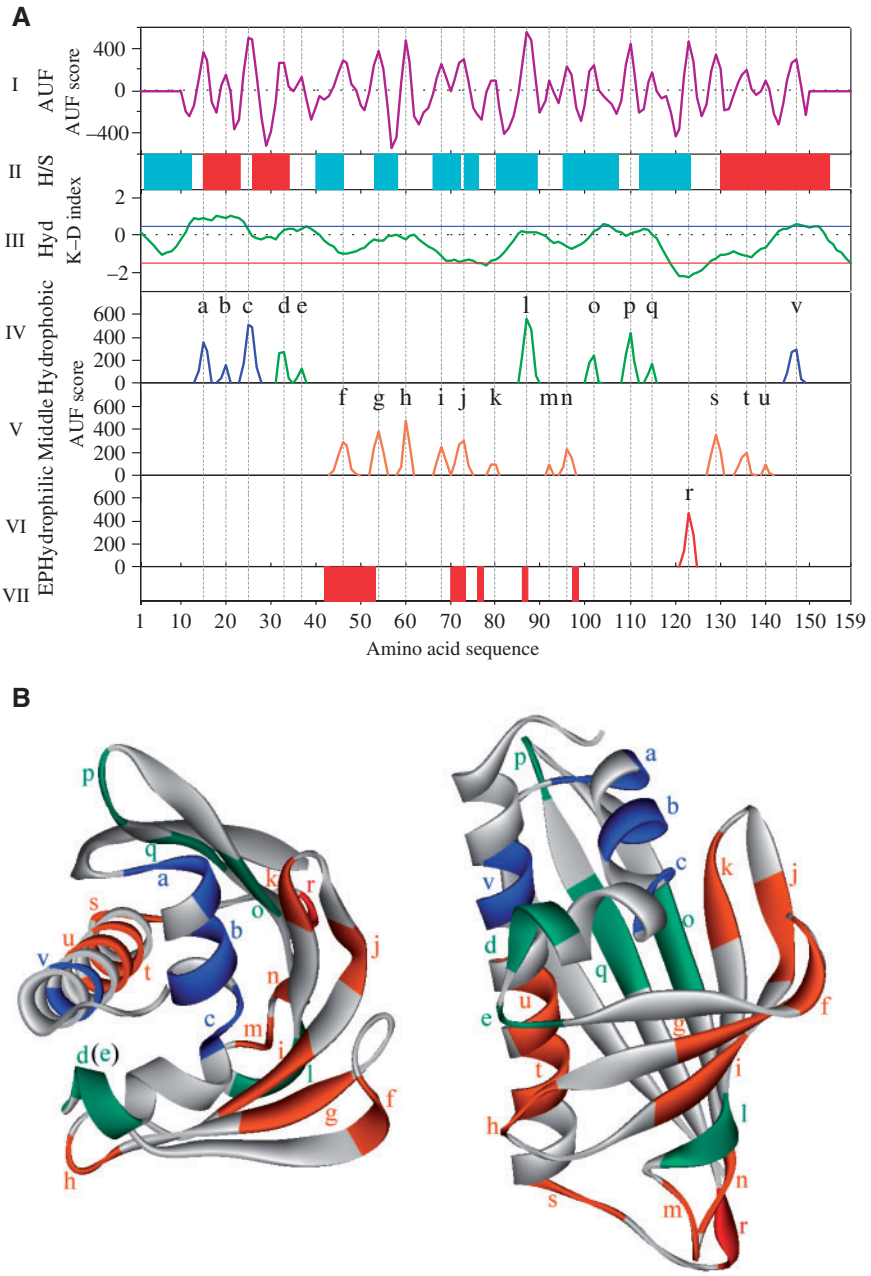

Fig. 4 AUF index plot in allergen Bet v. 1 with its related graphs (A) and its three-dimensional structure, in which the positions of the peaks in the AUF plot are coloured according to hydrophobicity (B). Graph I is the AUF index plot; II represents the secondary structures in which the red and blue regions indicate α-helixes and β-sheets, respectively; III is the hydrophobicity plot; IV shows the AUF peaks for $\langle H \rangle > 0.0$; V shows the AUF peaks for $-1.5 < \langle H \rangle < 0.0$: VI shows the AUF peaks for $\langle H \rangle < -1.5$, and VII shows the positions of the epitopes. The vertical dotted lines indicate the positions of the peaks of the AUF index plot.

peaks and binding sites suggests that the AUF plots will be a useful tool for predicting allergens.

Discussion

We analysed the amino acid distribution around the segments unique to allergens and developed the AUF plot. The utility of the AUF plot was then evaluated by examining the structural fluctuation of the birch pollen allergen, a typical allergen with known molecular structures in the single-molecule and complex states.

The fragments unique to allergen proteins were analysed and characteristic wavelet-like distributions of several amino acids were seen around the fragments. The shape of the distribution was largely symmetric, and the amino acids that are preferred by allergen epitopes showed peaks at the centre of the fragments and two shallow valleys on both sides of this peak. On the other hand, the amino acids avoided by epitopes showed sharp valleys with two low peaks on both sides of this valley. Analysing the distributions, we developed an index of amino acids for characterizing AUFs. Using this index together with the universal distribution function of amino acids around AUFs, a novel plot (AUF plot) was developed to represent the fragments similar to AUFs based on positive peaks.

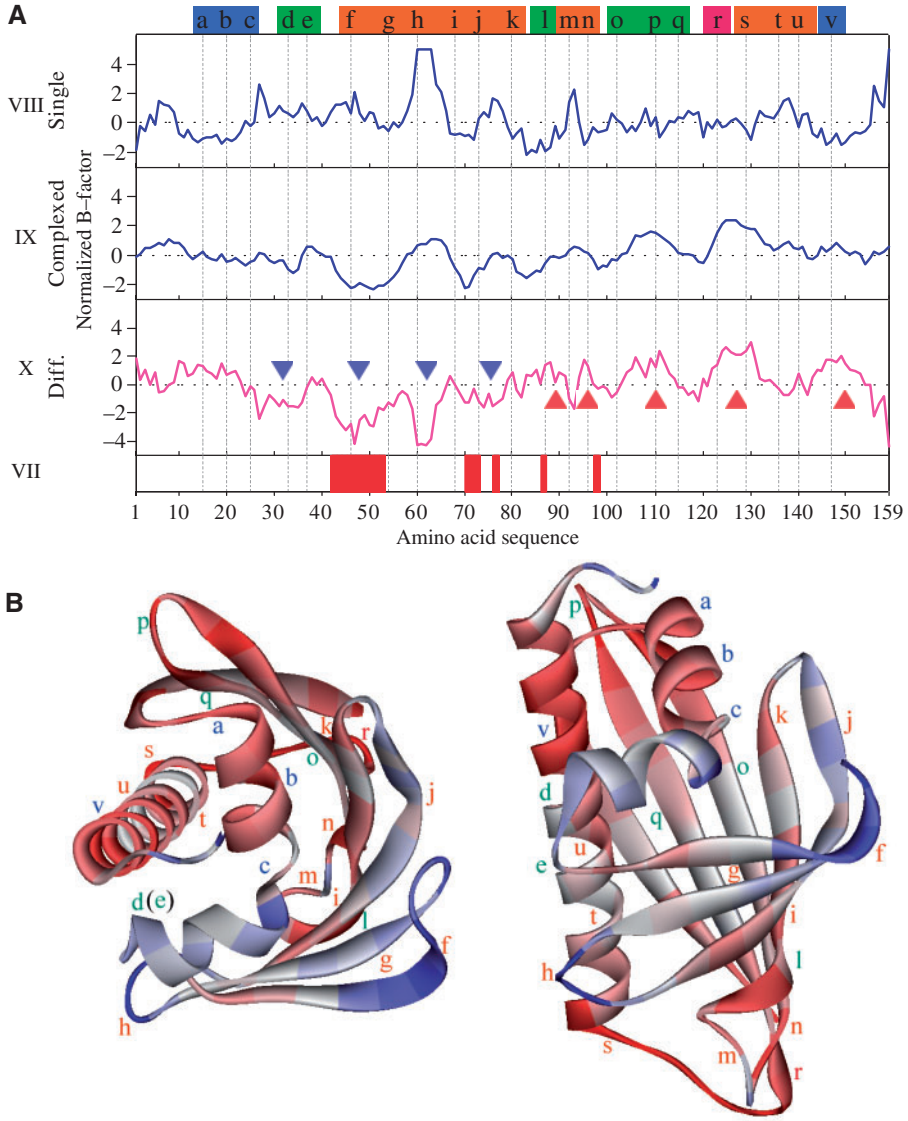

Fig. 5 (A) Plots of the B-factors of birch pollen allergen in the single-molecule state and in the complex with the corresponding IgE are shown in graphs VIII and IX, respectively. In graph X, the difference between graphs IX and VIII is plotted, in which two clusters in the regions with the downward (blue) and upward (red) arrows showed reduction and enhancement of structural fluctuation, respectively, based on IgE binding. Vertical dotted lines together with letters at the top of the graphs represent the peaks of the AUF index plot. (**B**) The molecular structure of the birch pollen allergen is coloured based on the difference in B-factor (graph X). Fluctuation is reduced by binding with the corresponding IgE, while the fluctuation in the other parts of the protein is enhanced.

Analysis of the amino acid sequences of birch pollen allergen showed that most peaks in the AUF plot were located at the loop regions or the end regions of secondary structures. Extensive analyses of numerous proteins showed the same tendency. Comparing the AUF plot obtained from the amino acid sequence with the degree of the fluctuation (B-factor) from the structural analysis of the birch pollen allergen, it was strongly suggested that the peaks of the AUF plot are closely related to the potential flexibility of amino acid fragments in proteins.

Many of the peaks in the AUF plots corresponded to B-factor peaks, thus suggesting that the amino acid sequences around the peaks of the AUF plots have large fluctuations in structure. The central regions of the peak have numerous flexible residues, such as glycine and alanine, as well as charged residues, which show high affinity to water, and this is consistent with the large fluctuation at the protein surface. Furthermore, the peripheral region contains numerous bulky aromatic residues and proline, which support large fluctuations in the central regions. It should be noted that these characteristics are not realized by high sequence homology. The conserved properties are not due to the sequences themselves, but rather their physical properties.

The AUF index appears to be useful in predicting the dynamic properties of fragments in the three-dimensional structure of proteins. However, the flexibility of a fragment is suppressed by binding with other molecules or domains, as shown in Fig. 5A. Therefore, some other filter that discriminates flexible segments

at the surface of a protein is necessary to predict dynamically fluctuating segments based on amino acid sequences alone. In fact, when segments at the peaks of the AUF plots are selected simply by average hydrophobicity, considerable discrimination in the flexible segments from segments at the core region with high AUF indexes is obtained. However, it is necessary for more accurate prediction to analyse the relationship between the peaks of AUF plots and the epitopes of many allergens, which will be described elsewhere.

References

1. Ragone, R., Facchiano, F., Facchiano, A., Facchiano, A.M., and Colonna, G. (1989) Flexibility plot of proteins. *Protein Eng*. **2**, 497–504
2. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., and Obradovic, Z. (2001) Intrinsically disordered protein. *J. Mol. Graph Model* **19**, 26–59
3. Demchenko, A.P. (2001) Recognition between flexible protein molecules: induced and assisted folding. *J. Mol. Recognit*. **14**, 42–61
4. Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry* **41**, 6573–6582
5. Li, X., Romero, P., Rani, M., Dunker, A.K., and Obradovic, Z. (1999) Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform Ser Workshop Genome Inform* **10**, 30–40
6. Linding, R., Russell, R.B., Neduva, V., and Gibson, T.J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res*. **31**, 3701–3708
7. Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol*. **337**, 635–645
8. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., and Dunker, A.K. (2003) Predicting intrinsic disorder from amino acid sequence. *Proteins* **53**, 566–572
9. Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003) Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453–1459
10. Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol*. **347**, 827–839
11. Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369–3376
12. Shimizu, K., Hirose, S., and Noguchi, T. (2007) POODLE-S: web application for predicting protein disorder by using physicochemical features and reduced amino acid set of a position-specific scoring matrix. *Bioinformatics* **23**, 2337–2338
13. Yuan, Z., Bailey, T.L., and Teasdale, R.D. (2005) Prediction of protein B-factor profiles. *Proteins* **58**, 905–912
14. Zorzet, A., Gustafsson, M., and Hammerling, U. (2002) Prediction of food protein allergenicity: a bioinformatic learning systems approach. *In Silico Biol*. **2**, 525–534
15. Fiers, M.W., Kleter, G.A., Nijland, H., Peijnenburg, A.A., Nap, J.P., and van Ham, R.C. (2004) Allermatch, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics* **5**, 133
16. Ivanciuc, O., Schein, C.H., and Braun, W. (2003) SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res*. **31**, 359–362
17. The Central Science Laboratory, The Allergen Database http://allergen.csl.gov.uk/ (October 27 2009, date last accessed)
18. Mills, C. Protall Project, Institute of Food Research, ProtAll http://www.ifr.ac.uk/Protall/ (October 27 2009, date last accessed)
19. Gajhede, M., Osmark, P., Poulsen, F.M., Ipsen, H., Larsen, J.N., Joost van Neerven, R.J., Schou, C., Løwenstein, H., and Spangfort, M.D. (1996) X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nat. Struct. Biol*. **3**, 1040–1045
20. Mirza, O., Henriksen, A., Ipsen, H., Larsen, J.N., Wissenbach, M., Spangfort, M.D., and Gajhede, M. (2000) Dominant epitopes and allergic cross-reactivity: complex formation between a Fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v 1. *J. Immunol*. **165**, 331–338
21. Spangfort, M.D., Mirza, O., Ipsen, H., Van Neerven, R.J., Gajhede, M., and Larsen, J.N. (2003) Dominating IgE-binding epitope of Bet v 1, the major allergen of birch pollen, characterized by X-ray crystallography and site-directed mutagenesis. *J. Immunol*. **171**, 3084–3090
22. Bannon, G.A. and Ogawa, T. (2006) Evaluation of available IgE-binding epitope data and its utility in bioinformatics. *Mol. Nutr. Food Res*. **50**, 638–644
23. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol*. **157**, 105–132
24. Imai, K. and Mitaku, S. (2005) Mechanisms of secondary structure breakers in soluble proteins. *BIOPHYSICS* **1**, 55–65
25. Smith, D.K., Radivojac, P., Obradovic, Z., Dunker, A.K., and Zhu, G. (2003) Improved amino acid flexibility parameters. *Protein Sci*. **12**, 1060–1072