ELISA

A competitive ELISA for quantitative determination of RBP4 in human plasma was performed according to the manufacturer's instructions (AdipoGen Inc.; Seoul, Korea). The detection limit was 1 ng/ml. An AHSG ELISA kit was used to detect fetuin-A in plasma (Bio-Vender Laboratory Medicine Inc.; Modrice, Czech Republic). The detection limit was 0.35 ng/ml. A Quantikine® Human Vitamin D-Binding Protein Immunoassay kit was used to detect VDBP in plasma (R&D Systems, Inc.; Minneapolis, MN, USA). The mean minimum detectable VDBP level was 0.65 ng/ml. Distribution of levels was represented using the median and interquartile range (IQR).

Statistical analysis

Proteins showing differential expression between two conditions were first determined with P values using the Student's t-test preinstalled in the PDQuest software suite. To select candidate proteins with expression levels that differed between unstimulated samples from patients with active TB and healthy control subjects, a significance level of $P < 0.05$ was selected. To select candidate proteins showing differential expression in *Mtb*-specific antigen-stimulated and unstimulated plasma samples, a less stringent cut-off value of $P < 0.10$ was applied. Assuming an alpha error of 0.1 and a standardized effect size of 2.0, the power to detect a difference was calculated as 0.8 given our sample size. When a normal distribution of measurements was not predicted, the Wilcoxon rank sum test (Mann-Whitney U test) was applied for confirmation using the JMP software (version 7.0.1; SAS Institute, Cary, NC, USA).

Results

Quantitative analyses by 2D-DIGE

In a preliminary experiment, we used an immobilized linear pH gel strip with a broad pH range (pH 3-10 linear) for 1D IEF. Although more than 500 protein spots were visualized in fresh plasma with SYPRO Ruby staining, the number of spots after incubation of whole blood with stimuli decreased, and detectable spots were primarily located in the pH range 4-7 (data not shown). Therefore, we performed subsequent analyses using an immobilized linear pH gel strip with a narrower range (pH 4-7 linear) to obtain a finer resolution. We used two comparative frameworks in our analyses, and the corresponding spot patterns are schematically depicted in Figure 1.

Differential gel images were acquired and displayed using the PDQuest 2D gel analysis software (Figure 2A, B). In our comparison of the protein expression profiles of patients with active TB and control subjects, red indicates proteins increased in the supernatants collected from the patients and green indicates proteins decreased in the patients compared with the control subjects. Yellow indicates no significant differences (Figure 2A). In 2D gel profiles comparing the antigen-stimulated and unstimulated samples collected from patients with active TB, red indicates proteins increased in the supernatants after *Mtb*-specific antigen stimulation, and green indicates proteins decreased after stimulation. Yellow indicates no significant changes (Figure 2B). From 367 spots compared between patients with active TB and control subjects, and 293 spots generated with samples collected from patients with active TB that were either stimulated with *Mtb*-specific antigens or left unstimulated, we selected several candidates for subsequent mass spectrometric analysis (Table 1) according to the criteria described in the Materials and Methods section.

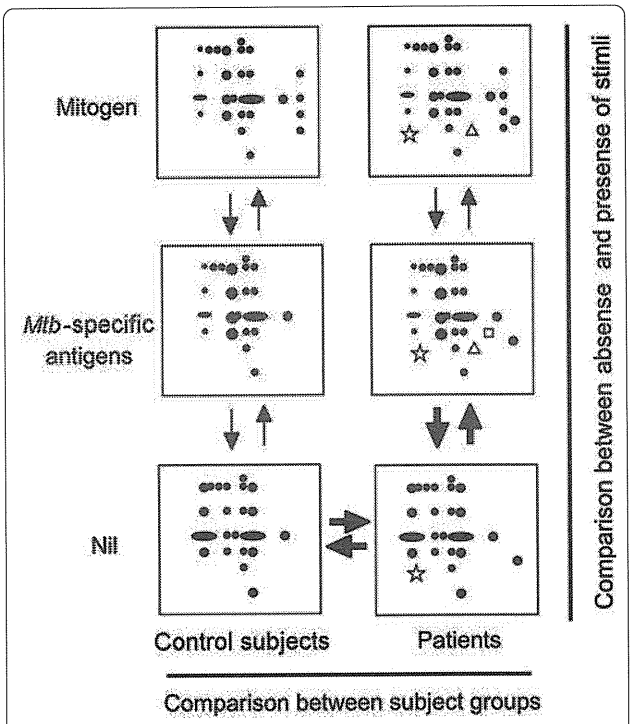


Figure 1 Schematic representation of protein expression profiles. Comparison between subjects, in particular, spot patterns showing differences in plasma proteomes between patients and control subjects. Comparison between the presence and absence of stimuli, especially *ex vivo* spot patterns of *Mtb*-specific antigen-stimulated plasma samples and unstimulated samples from patients. Two major comparative frameworks in our study are shown (bold arrows).

Mass spectrometric analysis

Following the above criteria for selecting candidates of differentially expressed proteins between two conditions, a total of 41 spots were isolated from the corresponding 2D gels on the basis that they showed sufficiently strong

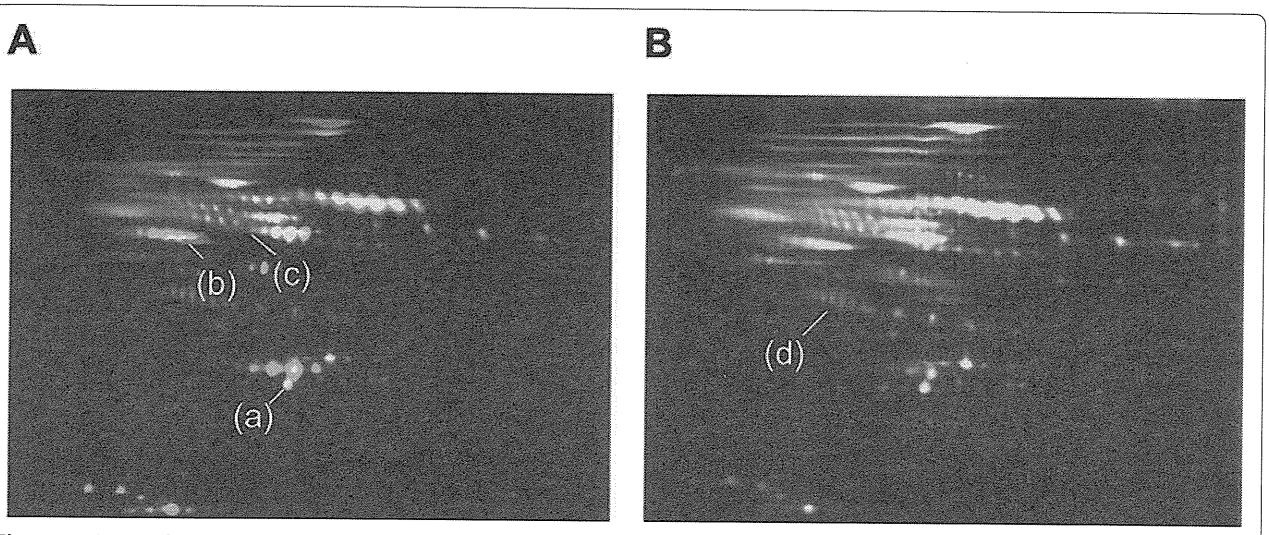


Figure 2 2D-DIGE-based protein expression analysis. Pseudo-colored images were generated using the Multi-Channel Viewer function of the PDQuest software suite. Four sample sets that were either unstimulated or stimulated with the *Mtb*-specific antigen or mitogen from patients with active pulmonary TB and four corresponding sets from control subjects were used to screen for candidate proteins by 2D-DIGE. A: Analysis based on differences in levels in samples from patients and control subjects. Red indicates proteins whose levels increased and green indicates proteins whose levels decreased in unstimulated supernatants from patients with active TB compared with control subjects. Spot (a): HT6102, Spot (b): HT2406, and Spot (c): HT5401 correspond to the PDQuest identification numbers shown in Table 2. B: Analysis based on differences in levels after stimulation. Red indicates proteins whose levels increased in whole blood supernatants from patients with active TB after incubation with *Mtb*-specific antigens. Spots (d): T3107, T4203, and T4208, correspond to Table 2.

signals. Trypsin digestion of each isolated spot was followed by LC-MS analysis. The proteins corresponding to 14 of these spots were successfully identified (Figure 3).

Of the 14 proteins in Table 2, 7 (serial numbers 1 to 7) were obtained as a result of comparisons between patients with active TB and control subjects; number 1 (spot HT6102) was identified as RBP4, number 2 (HT2406) as fetuin-A, and number 3 (HT5401) as VDBP. Four (numbers 8 to 11) were obtained as a result of comparisons between nonspecific mitogen-stimulated and unstimulated samples collected from patients with active TB (not analyzed in this study). The last 3 proteins (numbers 12 to 14) were obtained as a result of comparisons between *Mtb*-specific antigen-stimulated and unstimulated samples collected from patients with active TB; numbers 12 to 14 (T4203, T3107, and T4208) were all identified as clusterin. In Table 2, *P* values indicating a significant difference between the means of the two conditions examined, the SWISS-PROT accession numbers of the identified proteins as well as their molecular weights and theoretical pI values are indicated. We also used the *Homo sapiens* database of expressed sequence tags (ESTs) to identify clusterin in spot T4208.

Mascot search scores (indices of protein matches) were 47, 50, 98, 75, and 72 for spots T4203 (clusterin), T3107(clusterin), HT5401 (VDBP), HT2406 (fetuin-A), and HT6102 (RBP4), respectively, (Table 2), suggesting that identification of these proteins using peak lists of MS/MS spectra obtained from the LC-MS/MS system are fairly reliable since all these scores were significant

Table 1 The number of spots that may show differential expression

A: Comparison between patients with active TB and control subjects				
		***P* < 0.02**	**0.02≤*P* < *0.05***	**0.05≤*P* < 0.10**
Patients versus control subjects		18	12	24
B: Comparison between stimulated and unstimulated conditions				
		P < 0.02	**0.02≤P < 0.05**	**0.05≤P < 0.10**
Patients	*Mtb* antigens versus no stimuli	0	2	2
	Mitogen versus no stimuli	3	5	11
Control subjects	*Mtb* antigens versus no stimuli	0	1	13
	Mitogen versus no stimuli	2	83	8

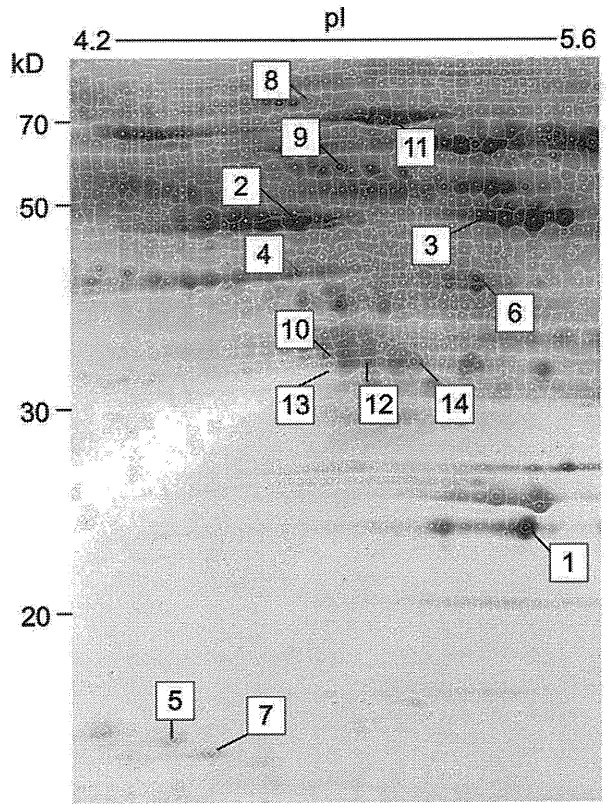


Figure 3 Representative 2D gel image obtained from pooled protein samples. Differentially expressed proteins successfully identified by LC-MS were represented on the gel using the PDQuest software suite. See Table 2 for additional information. Numbers in the figure correspond to serial numbers in Table 2.

above the 5% confidence threshold and no other proteins with comparable scores were detected for each gel spot (See Additional file 1: for supporting information). These proteins were interesting because of their potential biological significance, and we therefore analyzed them further.

Confirmation of differentially expressed proteins by immunoblotting

Immunoblot analysis was used to confirm differential expression of three proteins identified in patients with active TB compared with control subjects (Figure 4A). We measured band densities using the same samples prepared for protein confirmation (Figure 4B). The band density of RBP4 in patients with active TB (64,283 arbitrary units ± 3,861) was lower than that in control subjects (445,894 ± 16,590), and fetuin-A expression in the patients was also lower (42,710 ± 7,580) than that in control subjects (343,617 ± 58,923). These results are consistent with those of 2D gel analysis. Moreover, the band density of VDBP tended to be higher in samples from patients with active TB than from control subjects, which is similar to that observed above; however, the protein levels were widely distributed and the differences in these levels did not reach significance in the control subjects compared with patients with active TB (33,251 ± 2,572 versus 38,971 ± 11,001). Because the three clusterin spots altered after *Mtb*-specific antigen stimulation were not clearly distinguished by immunoblotting, we did not attempt any further demonstration of changes in these signals in our study. Instead, pooled samples were run on a 2D gel and followed by immunoblotting with anti-clusterin-α and anti-clusterin-β antibodies (because clusterin consists of clusterin-α and -β subunits) (Figure 4C). Based on immunoreactivity and pI values, the spots detected were confirmed to be clusterin-α. More specifically, the three spots comprised a subset of possible modified forms of clusterin-α that may be detected.

Detection of differentially expressed proteins by ELISA

Because RBP4 and fetuin-A levels determined by immunoblotting were significantly different between samples from patients with active TB and control subjects, we performed further quantitative ELISA to extend the measurements to plasma samples from 14 Japanese patients with active TB and 13 age-, gender-, and ethnicity-matched control subjects. Plasma RBP4 levels in patients with active TB (median = 23.6 μg/ml; IQR = 18.4-37.9) were significantly lower than those from control subjects (median = 44.6 μg/ml; IQR = 34.6-53.8; P = 0.0033; Figure 5A). Plasma fetuin-A levels in patients (median = 147.9 μg/ml; IQR = 115.8-159.6) were also significantly lower than those in control subjects (median = 211.0 μg/ml; IQR = 186.7-264.6; P = 0.0002; Figure 5B). No significant difference were observed in plasma VDBP levels between patients (median = 110.0 μg/ml; IQR = 85.2-151.3) and control subjects (median = 105.0 μg/ml; IQR = 88.1-215.6; P = 0.5441; Figure not shown).

We simultaneously compared the protein levels in plasma immediately separated from EDTA-containing blood with those in plasma supernatants obtained from heparinized blood as a negative control for IGRA after 18 h of incubation without stimulants. We found that the differences between the two types of plasma samples were small (coefficient of variance (CV) = 10.5% for RBP4; CV = 5.0% for fetuin-A; CV = 6.6% for VDBP) and was in a range of variation generally accepted in ELISA (CV < 15%), indicating that the measurements obtained under the latter condition can be substituted for those obtained under the former condition. Indeed, plasma RBP4 and fetuin-A levels in samples from Japanese patients with active TB were significantly lower than those from control subjects, irrespective of plasma conditions (data not shown).

Table 2 Characteristics of proteins identified in this study

A: Comparison between patients with active TB and control subjects

		2D-DIGE			LC-MS/MS				
Condition	**Serial Number**	**PDQ SSP#[a]**	***P* value**	**+/-[b]**	**Swiss-Plot**	**Mascot search score[e]**	**Protein name**	**Da**	**pI**
Patients versus control subjects	1	HT6102	0.0064	-	RET4_HUMAN	72	Retinol binding protein 4	23010	5.76
	2	HT2406	0.0097	-	FETUA_HUMAN	75	α-2-HS-glycoprotein	39325	5.43
	3	HT5401	0.0331	+	VTDB_HUMAN	98	Vitamin D binding protein	52964	5.40
	4	HT2303	0.0419	+	CO4A_HUMAN	86	Complement C4A	192771	6.66
	5	HT1012	0.0271	-	APOC3_HUMAN	105	Apolipoprotein C-III	10852	5.23
	6	HT5303	< 0.001	-	APOA4_HUMAN	190	Apolipoprotein A-IV	45399	5.28
	7	HT1016	0.0024	-	APOC2_HUMAN	61	Apolipoprotein C-II	11284	4.72

B: Comparison between stimulated and unstimulated conditions in active TB

		2D-DIGE			LC-MS/MS				
Condition	**Serial Number**	**PDQ SSP#**	***P* value**	**+/-[c]**	**Swiss-Plot**	**Mascot search score**	**Protein name**	**Da**	**pI**
Mitogen (versus no stimuli)	8	T3601	0.0917	-	C1S_HUMAN	169	Complement-C1S	76684	4.86
	9	T3403	0.0156	+	KNG1_HUMAN	139	Kininogen-1	71957	6.34
	10	T3105	0.0866	-	ZA2G_HUMAN	45	Zinc-α-2-glycoprotein	33872	5.57
	11	T4512	0.0061	-	A1BG_HUMAN	76	α-1B-glycoprotein	54273	5.58
Mtb antigens (versus no stimuli)	12	T4203	0.0640[d]	+	CLUS_HUMAN	47	Clusterin	52495	5.89
	13	T3107	0.0687[d]	+	CLUS_HUMAN	50	Clusterin	52495	5.89
	14	T4208	0.0732[d]	+	EST	–	Clusterin	52495	5.89

[a]PDQ SSP# is a PDQuest software standard spot number indicating the unique location of each spot automatically assigned on a 2D gel and is essential for comparing the same spots on different gels.

[b]The average density of a spot from 2D-DIGE is higher (+) or lower (-) in patients with active TB than in control subjects.

[c]The average density of the spot from 2D-DIGE is higher (+) or lower (-)under the stimulated condition than that under the unstimulated condition.

[d]The average density of the 3 spots, T4203, T3107, and T4208, which correspond to a subset of clusterin, was significantly higher in *Mtb*-specific antigen-stimulated than in unstimulated samples ($P = 0.0014$).

[e]The Mascot search score indicates the degree of compatibility between mass spectra generated by the sample and amino acid sequences within the protein of interest.

We further attempted to verify the differences observed with samples from a different ethnic and regional population, i.e., samples collected from Vietnamese patients. The two proteins identified above were measured in plasma supernatants from Vietnamese patients with active TB and age-, gender-, and ethnicity-matched control subjects. The samples from these Vietnamese patients were obtained from a negative control of IGRA after incubation without stimulants. RBP4 levels in patients with active TB (median = 17.5 μg/ml; IQR = 14.4-23.9) were significantly lower than those in control subjects (median = 30.5 μg/ml; IQR = 25.9-40.8; $P < 0.0001$; Figure 5A). Fetuin-A levels in patients with active TB (median = 210.7 μg/ml; IQR = 178.1-235.7) were also significantly lower than those in control subjects (median = 299.4 μg/ml; IQR = 265.1-363.2; $P < 0.0001$; Figure 5B). Moreover, both protein levels were not significantly different between IGRA-negative and IGRA-positive subgroups of the control subjects (data not shown).

Discussion

In this study, we identified TB-associated proteins from whole blood supernatants. After the removal of 14 major plasma proteins, RBP4, fetuin-A, and VDBP were initially identified as plasma proteins from unstimulated samples for which the baseline levels differed between the patients and control subjects. Immunoblotting results confirmed the differential expression of RBP4 and fetuin-A between the two groups. Although VDBP has previously been identified as a biomarker for mycobacterial infections in cattle [9], the level of this protein did not differ significantly in our study because of large individual variations. The changes in VDBP levels may not have been accurately immunologically assayed in our study.

Clusterin is a secreted glycoprotein involved in apoptosis, inflammation, and tissue injury. It was differentially expressed in patients with active TB after stimulation and the intensities of the three spots

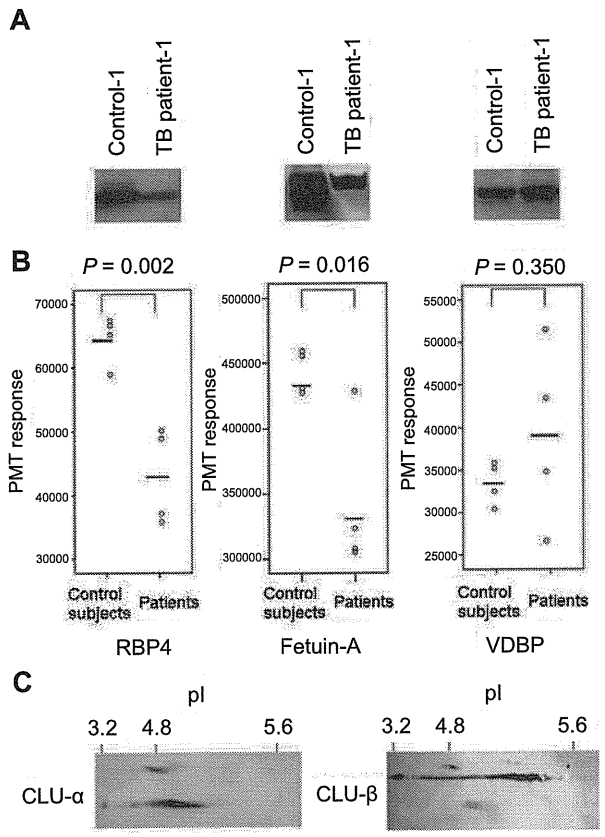


Figure 4 Comparison of candidate differential proteins using immunoblotting. Samples were first treated to remove 14 major proteins and then analyzed by immunoblotting to facilitate comparison of RBP4, fetuin-A, and VDBP levels. (A) Representative immunoblot of samples from patient with active TB (Patient-1) versus a control subject (Control-1). (B) Band densities were quantified in patients (n = 4) and control subjects (n = 4). The PMT response suggested band densities. (C) Two-dimensional immunoblot of pooled samples with detection of clusterin-α and -β. CLU is clusterin.

corresponding to clusterin-α were elevated in whole blood supernatant samples after incubation with *Mtb*-specific antigens. These spots appear to have shifted in both the dimensions on the gel, which suggests small changes in their molecular weights and IEPs. It is conceivable that post-translational modifications, such as degradation and/or deglycosylation, occur via an enzymatic reaction that accompanies immune cell activation. However, we have not demonstrated that this response is observed only when *Mtb*-specific antigens are co-incubated. To determine whether clusterin has a role as a marker of TB or indicates more general response to antigen stimulation, we are currently attempting to find clear and simple methods in detecting these alterations for mass screening.

Subsequent ELISA results for samples from Japanese and Vietnamese subjects confirmed that both plasma RBP4 and fetuin-A levels were significantly lower in samples collected from patients with active TB than in control subjects, indicating that our findings are reproducible in studies using well-matched control subjects. However, as shown in Figure 5, the average plasma levels of these proteins differed between Japanese and Vietnamese control subjects. This suggests that unknown factors may systemically influence tested populations or the measurement of these markers. Because this variance is crucial in a clinical setting, further basic as well as clinical investigations are necessary to accurately assess these markers.

No significant differences were observed in RBP4 and fetuin-A levels in samples from IGRA-positive and -negative control subjects. This suggests that these proteins levels are not affected by latent TB infection, but that they presumably change during disease progression via an unknown mechanism.

Intriguingly, the literature supports the idea that RBP4 and fetuin-A are functionally significant since they may be involved in macrophage activation [10-12]. Retinoic acid has been shown to stimulate and induce monocyte differentiation, leading to inhibition of *Mtb* multiplication in human macrophages [13]. RBP4 is the specific carrier protein for retinol (vitamin A) and has recently been described as an adipokine that contributes to insulin resistance [13]. This protein is believed to modulate pathophysiological processes during bacterial infection. Fetuin-A was originally identified as a fetal protein and has been shown to affect the development of many mammalian tissues. Moreover, the results of *in situ* mRNA hybridization and immunocytochemical studies in adult sheep have revealed that the main sites of fetuin-A expression are hepatocytes and monocytes or macrophages in the spleen and bone marrow [14]. Fetuin-A is known to modulate various immune and metabolic responses. Previous reports have shown that fetuin-A deactivates macrophages, acts as an opsonin for cationic-deactivating molecules including spermine [15], reduces TNF-α production and inflammatory responses [16], and enhances phagocytosis of apoptotic cells and macropinocytosis by human macrophages [17]. On the other hand, this protein is known to be a potent inhibitor of systemic calcification [18] and is associated with the incidence of diabetes mellitus [19].

Our study is the first to highlight the relationship between these two markers and TB, even though these marker levels may be affected by endogenous or exogenous factors and are presumably nonspecific to TB given their relative abundance in plasma and the broad spectrum of functional significance proposed in the above references.

Nevertheless, performing a prospective cohort study may help clarify the role of these proteins in TB.

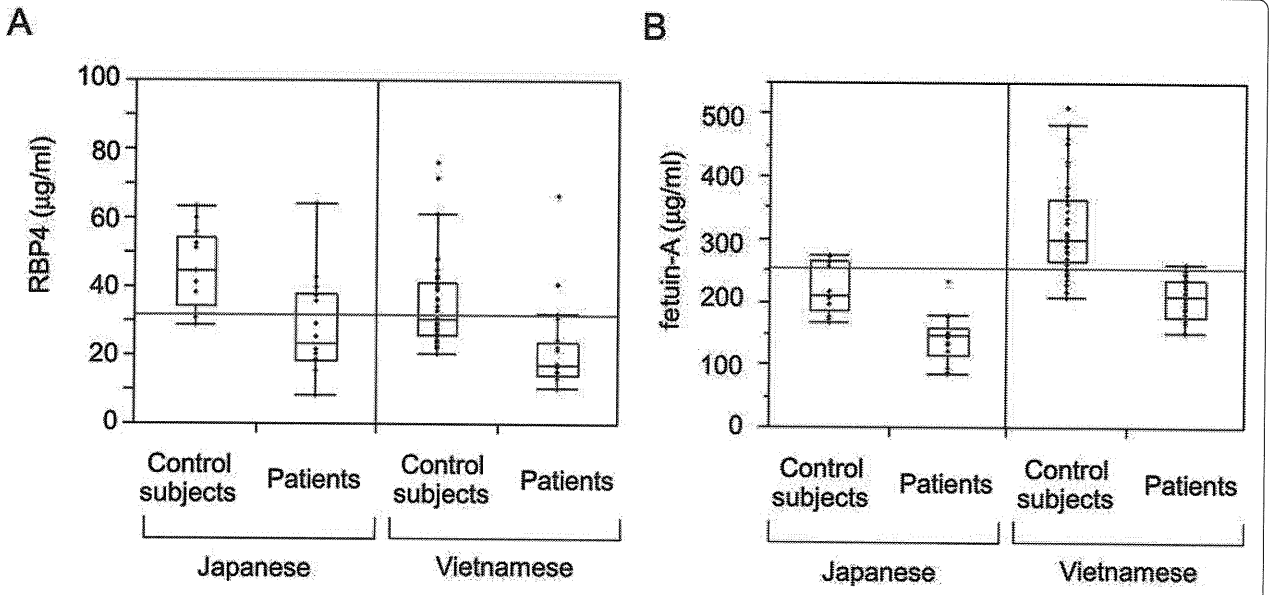


Figure 5 Comparison of RBP4, fetuin-A, and VDBP levels by ELISA. Samples were analyzed by ELISA to compare RBP4 and fetuin-A levels in14 Japanese patients with active TB and 13 age-, gender-, and ethnicity-matched control subjects, as well as 25 Vietnamese patients with active TB, and 50 age-, gender-, and ethnicity-matched control subjects. *P* values showing results of comparisons are described in the text. (A) RBP4 levels in patients versus control subjects. (B) Fetuin-A levels in patients versus control subjects.

If within-individual variation in baseline levels is relatively small, it can be used to monitor the course of disease before, during, and after treatment. Further clinical studies on various conditions may better characterize these proteins. Single use of these markers or their combined use with other promising biomarkers may be a useful tool to aid the development of new effective therapies and vaccines.

Conclusions

We identified three TB-associated proteins, RBP4, fetuin-A, and clusterin, in whole blood supernatants using a proteomic approach. We subsequently showed that both plasma RBP4 and fetuin-A levels are significantly and reproducibly lower in patients with active TB than in control subjects. These findings may help us understand and monitor the disease process in TB.

Additional material

Additional file 1: Figure S1 - Mascot search results–Information about the identified proteins obtained using the Mascot server. (A) Mascot search result for T2116 (clusterin) (B) Mascot Search Result for T2103 (clusterin) EST (C) Mascot search result for T1486 (clusterin) (D) Mascot search result for HT2482 (RET4 = RBP4) (E) Mascot search result for HT1248 (fetuin-A) (F) Mascot search result for HT1240 (VDBP).

Abbreviations

TB: Tuberculosis; *Mtb*: *Mycobacterium tuberculosis*; IFN-γ: interferon-gamma; IRGA: interferon-gamma-release assay; 2D-DIGE: two-dimensional difference gel electrophoresis; LC-MS: liquid chromatography-mass spectrometry; ELISA: enzyme-linked immunosolvent assay; IEF: isoelectric focusing; RBP-4: retinol binding protein-4; VDBP: vitamin D binding protein; IQR: interquartile range; ESTs: expressed sequence tags;

Acknowledgements

The authors thank Dr. Harada of the Research Institute of Tuberculosis; Drs. Nakamichi and Hanada of the Department of Respiratory medicine, National Center for Global Health and Medicine; and T. Totsu and Dr. Kanazawa of the Department of Respiratory Diseases, Research Institute, National Center for Global Health and Medicine for their assistance with the project. This work was partly supported by a grant of National Center for Global Health and Medicine and by a grant from the Program of Japan Initiative for Global Research Network on Infectious Diseases (J-GRID), MEXT, Japan.

Author details

[1]Molecular Enzymology, Graduate School of Agricultural Science, Tohoku University, 1-1 Amamiya-machi, Tsutsumidori, Aoba-ku, Sendai, Miyagi 981-8555, Japan. [2]Department of Respiratory Diseases, Research Institute, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan. [3]Department of Diabetic Complications, Research Institute, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan. [4]Department of Metabolic Disorder, Research Institute, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan. [5]Supramolecular Biology, International Graduate School of Art and Sciences, Yokohama City University, 1-7-29 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [6]Department of Respiratory Medicine, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan. [7]National Center for Global Health and Medicine - Bach Mai Hospital (NCGM-BMH) Medical Collaboration Center, 78 Giai Phong St., Hanoi, Vietnam. [8]Hanoi Tuberculosis and Lung Disease Hospital, 44 Thanh Nhan Road, Hanoi, Vietnam.

Authors' contributions

TT carried out the plasma proteome studies, participated in 2D-DIGE studies, a part of immunoassays and drafted the manuscript. SS conceived of the study, and participated in the study planning and coordination and helped to draft the manuscript. KK carried out the LC-MS/MS analysis. ET, KY and

HH helped to design the study. YK participated in the study design and overall supervision. NK, NTLH and LTL participated in management and analysis of data. IM and MH participated in the acquisition of data. TU helped to draft the manuscript. NK participated in the design of the study, performed statistical analysis and have given final approval of the version to be published. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 7 October 2010 Accepted: 22 March 2011
Published: 22 March 2011

References

1. WHO: **Global Tuberculosis Control: Epidemiology, Strategy, Financing. Geneva.** WHO Press; 2009, 7-13.
2. Jacobsen M, Repsilber D, Guschmidt A, *et al*: **Candidate biomarkers for discrimination between infection and disease caused by Mycobacterium tuberculosis.** *J Mol Med* 2007, **85**:613-621.
3. Mistry R, Cliff JM, Clayton CL, *et al*: **Gene expression patterns in whole blood identify subjects at risk for recurrent tuberculosis.** *J Infect Dis* 2007, **195**:357-365.
4. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, Loosemore A, Tarelli E, Sheldon J, Schwenk A, Pollok R, Rayner CF, Krishna S: **Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum.** *Lancet* 2006, **368**:1012-1021.
5. Chegou NN, Black GF, Kidd M, van Helden PD, Walzl G: **Host markers in Quantiferon supernatants differentiate active TB from latent TB infection: preliminary report.** *BMC Pulm Med* 2009, **9**:21.
6. Wallis RS, Pai M, Menzies D, Doherty TM, Walzl G, Perkins MD, Zumla A: **Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice.** *Lancet* 2010, **375**:1920-1937.
7. Toda T, Sugimoto M, Omori S, Matsuzaki T, Furuichi Y, Kimura N: **Proteomic analysis of Epstein-Barr virus-transformed human B-lymphoblastoid cell lines before and after immortalization.** *Electrophoresis* 2000, **21**:1814-1822.
8. Saeki K, Yasugi E, Okuma E, Breit SN, Nakamura M, Toda T, Kaburagi Y, Yuo A: **Proteomic analysis on insulin signaling in human hematopoietic cells: identification of CLIC1 and SRp20 as novel downstream effectors of insulin.** *Am J Physiol Endocrinol Metab* 2005, **289**:E419-428.
9. Seth M, Lamont EA, Janagama HK, Widdel A, Vulchanova L, Stabel JR, Waters WR, Palmer MV, Sreevatsan S: **Biomarker discovery in subclinical mycobacterial infections of cattle.** *PLoS One* 2009, **4**:e5478.
10. Crowle AJ, Ross EJ: **Inhibition by retinoic acid of multiplication of virulent tubercle bacilli in cultured human macrophages.** *Infect Immun* 1989, **57**:840-844.
11. Jersmann HP, Dransfield I, Hart SP: **Fetuin/alpha2-HS glycoprotein enhances phagocytosis of apoptotic cells and macropinocytosis by human macrophages.** *Clin Sci* 2003, **105**:273-278.
12. Wang H, Zhang M, Bianchi M, Sherry B, Sama A, Tracey KJ: **Fetuin (alpha2-HS-glycoprotein) opsonizes cationic macrophagedeactivating molecules.** *Proc Natl Acad Sci USA* 1998, **95**:14429-14434.
13. Yang Q, Graham TE, Mody N, Preitner F, Peroni OD, Zabolotny JM, Kotani K, Quadro L, Kahn BB: **Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes.** *Nature* 2005, **436**:356-362.
14. Dziegielewska KM, Brown WM, Deal A, Foster KA, Fry EJ, Saunders NR: **The expression of fetuin in the development and maturation of the hemopoietic and immune systems.** *Histochem Cell Biol* 1996, **106**:319-330.
15. Wang H, Zhang M, Bianchi M, Sherry B, Sama A, Tracey KJ: **Fetuin (α2-HS-glycoprotein) opsonizes cationic macrophagedeactivating molecules.** *Proc Natl Acad Sci USA* 1998, **95**:14429-14434.
16. Ombrellino M, Wang H, Yang H, Zhang M, Vishnubhakat J, Frazier A, Scher LA, Friedman SG, Tracey KJ: **Fetuin, a negative acute phase protein, attenuates TNF synthesis and the innate inflammatory response to carrageenan.** *Shock* 2001, **15**:181-185.
17. Jersmann HP, Dransfield I, Hart SP: **Fetuin/α2-HS glycoprotein enhances phagocytosis of apoptotic cells and macropinocytosis by human macrophages.** *Clin Sci (Lond)* 2003, **105**:273-278.
18. Jahnen-Dechent W, Schafer C, Ketteler M, McKee MD: **Mineral chaperones: a role for fetuin-A and osteopontin in the inhibition and regression of pathologic calcification.** *J Mol Med* 2008, **86**:379-389.
19. Ix JH, Wassel CL, Kanaya AM, Vittinghoff E, Johnson KC, Koster A, Cauley JA, Harris TB, Cummings SR, Shilipak MG: **Fetuin-A and incident diabetis mellitus in older persons.** *JAMA* 2008, **300**:182-188.

Pre-publication history
The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2334/11/71/prepub

doi:10.1186/1471-2334-11-71
Cite this article as: Tanaka *et al.*: **Identification of tuberculosis-associated proteins in whole blood supernatant.** *BMC Infectious Diseases* 2011 **11**:71.

Genet Resour Crop Evol (2011) 58:283–289
DOI 10.1007/s10722-010-9573-5

RESEARCH ARTICLE

Genetic variation of high-molecular-weight glutenin subunit composition in Asian wheat

Yohei Terasawa · Kanenori Takata · Hisashi Hirano · Kenji Kato · Taihachi Kawahara · Tetsuo Sasakuma · Tsuneo Sasanuma

Received: 28 December 2009 / Accepted: 10 May 2010 / Published online: 11 June 2010

Abstract To clarify the genetic properties of the HMW glutenin subunit composition of Asian endemic wheats, SDS–PAGE analysis was conducted using 1,139 bread wheat accessions that were originally collected in Asia. The samples were divided into six regional groups, Western Asia, Caucasia, Central Asia, Afghanistan, Southern Asia, and Eastern Asia. The genotype *Glu-A1c*, *Glu-B1b*, and *Glu-D1a* encoding subunits null, 7+8, and 2+12 had an overall frequency of 55.2%. Thus, we conclude that it is the typical genotype of the HMW glutenin subunits that characterize Asian endemic wheat. The frequency of the typical Asian genotype was relatively high in the central belt of Asia (Western Asia, Afghanistan, and Eastern Asia) and low in the marginal regions (Caucasia, Central Asia, and Southern Asia). In Southern Asia, the frequency of *Glu-B1i*, which encodes subunit 17+18, was the highest at the *Glu-B1* locus. In Caucasia and Central Asia, the frequency of *Glu-D1d*, which encodes subunit 5+10 (which is considered to be the most useful for making bread), was high. The level of genetic variation, as estimated using the frequencies of the various alleles, was relatively low in the central belt of Asia and high in the marginal regions. Among the three *Glu-1* loci, the highest number of alleles was detected at the *Glu-D1* locus. This result was caused by the presence of rare Asian specific alleles at the *Glu-D1* locus, in which a newly found allele, *Glu-D1bs*, encoding subunit 2.1+12 was included.

Keywords Asian endemic wheat · Genetic variation · *Glu-1* loci · HMW-glutenin subunit · *Triticum aestivum*

Y. Terasawa · T. Sasakuma · T. Sasanuma
Kihara Institute for Biological Research, Yokohama City University, Maioka-cho 641-12, Totsuka-ku, Yokohama 244-0813, Japan

Y. Terasawa · H. Hirano
International Graduate School of Arts and Science, Yokohama City University, Seto 22-2, Kanazawa-ku, Yokohama 236-0027, Japan

Y. Terasawa · T. Sasanuma (✉)
Faculty of Agriculture, Yamagata University, Wakaba-machi 1-23, Tsuruoka, Yamagata 997-8555, Japan
e-mail: sasanuma@tds1.tr.yamagata-u.ac.jp

K. Takata
National Agricultural Research Center for Western Region, Fukuyama, Hiroshima 721-8514, Japan

K. Kato
Graduate School of Natural Science and Technology, Okayama University, 3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, Japan

T. Kawahara
Laboratory of Crop Evolution, Plant Germ-Plasm Institute, Graduate School of Agriculture, Kyoto University, Mozume, Muko, Kyoto 617-0001, Japan

Introduction

The high-molecular-weight (HMW) glutenin subunit is one of the major components of the wheat seed

storage protein, the composition of which affects flour quality (Shewry et al. 1992). This protein is encoded by three homoeologous loci located on group 1 chromosomes, *Glu-A1*, *Glu-B1*, and *Glu-D1*, among which the loci on the B and D genomes encode duplicated genes, which are named the x- and y-types (Payne et al. 1980, 1982). Due to its agronomical importance, the genetic variation of the HMW glutenin subunit has been studied in numerous bread wheats (*Triticum aestivum* L.), and various kinds of subunits and alleles encoding these subunits have been identified (Payne and Lawrence 1983; Lee et al. 1995; Wrigley et al. online).

Since Asia has a long history of wheat cultivation, a number of endemic varieties have arisen in Asia, which are adapted to the topography, climate, and culture of each region. Hence, it is expected that a high level of genetic variation and unique genotypes are maintained in Asian wheat. To broaden the genetic resources available for flour breeding, HMW glutenin subunit composition has been extensively studied in wheats endemic to several Asian countries, e.g. Japan (Nakamura 2000), China (Zhang et al. 2002; Cong et al. 2005; Fang et al. 2009), and Pakistan (Tahir et al. 1996). Recently, we investigated the genetic variation of HMW glutenin subunit composition in Afghan wheat landraces (Terasawa et al. 2009). Although Afghanistan is regarded as one of the centers of crop diversity (Vavilov 1926), we found that Afghan wheat had a relatively low level of HMW glutenin subunit composition diversity compared to wheats from neighboring countries. One genotype, *Glu-A1c*, *Glu-B1b*, and *Glu-D1a*, which encodes subunits null, 7+8, and 2+12, was present in more than 80% of Afghan wheat. Interestingly, it was reported that this genotype is also the most common genotype in Japan and China (Nakamura 2000; Cong et al. 2005). Thus, to comprehensively understand the genetic properties of HMW glutenin subunit composition in Asia, further study using a greater number of Asian wheats collected from other regions is necessary.

In this study, we investigated the genetic variation of HMW glutenin subunit composition in more than one thousand wheat accessions collected from the center of Asia. Based on the results, the characteristics of HMW glutenin subunit composition in Asian endemic wheats are discussed.

Materials and methods

Plant materials

A total of 1,139 accessions of bread wheats endemic to Asia were used in this study. These samples were originally collected in 17 Asian countries and regions. The samples were divided into six regional groups: Caucasia (including Azerbaijan, Armenia, and Georgia), Central Asia (Tajikistan, Turkmenistan, Uzbekistan, Kyrgyzstan, and Kazakhstan), Western Asia (Iraq and Iran), Afghanistan, Southern Asia (Pakistan, India, Nepal, and Bhutan), and Eastern Asia (two regions in China) (Table 1). Among them, 1,017 accessions were from the collection of the Plant Germplasm Institute, Kyoto University, or the Kihara Institute for Biological Research, Yokohama City University, and were provided through the National BioResources Project (NBRP) KOMUGI. Ninety-eight accessions were donated by USDA, and 24 accessions belonged to one of the authors (K. Kato).

SDS polyacrylamide gel electrophoresis (SDS–PAGE) analysis

HMW glutenin subunit composition was determined by SDS–PAGE according to the method of Terasawa et al. (2009). The data for all the accessions of Afghanistan and some of the accessions of Iran and Pakistan were derived from our previous paper (Terasawa et al. 2009). To evaluate genetic diversity, the polymorphic information content index (PIC), which is equivalent to Nei's diversity index (Nei 1973), was calculated based on the frequencies of the alleles at each *Glu-1* locus and the allelic combinations of the loci.

Results

Characterization of HMW glutenin subunit composition

The allelic frequencies at each *Glu-1* locus are shown in Table 2. For the *Glu-A1* locus, three alleles were found, among which the *Glu-A1c* allele was the most frequent, both overall (74.4%) and in each of the regions. The second most common allele was *Glu-A1b*.

Table 1 Number of accessions used in this study

Region	No. of accessions	Country or internal region	No. of accessions
Western Asia	68	Iraq	19
		Iran[a]	49
Caucasia	26	Azerbijian	3
		Armenia	13
		Georgia	10
Central Asia	81	Tajikistan	18
		Turkmenistan	14
		Uzbekistan	19
		Kyrgyzstan	14
		Kazakhstan	16
Afghanistan	410	Afghanistan[a]	410
Southern Asia	93	Pakistan[a]	48
		India	25
		Nepal	10
		Bhutan	10
Eastern Asia	461	Xinjiang Uygur in China	124
		Sichuan and Tibet in China	337
Total	1,139		

[a] All the accessions of Afghanistan, 33 accessions of Iran, and 32 accessions of Pakistan were used in our previous study (Terasawa et al. 2009)

This allele was especially common in Caucasia and Central Asia, in which it had a frequency of more than 30%.

At the *Glu-B1* locus, nine alleles were found, among which *Glu-B1b* occurred in 76.5% of the accessions. Although this allele was the most common in all regions except for Southern Asia, its frequency was relatively low in Caucasia (42.3%) and Central Asia (60.5%). In Southern Asia, the frequency of *Glu-B1b* was 38.9%, and the most frequent allele was *Glu-B1i* (46.7%), which was a minor allele in other regions. Regarding other alleles, a unique feature was observed in Caucasia and Central Asia. In these regions, the frequency of *Glu-B1a* (34.6 and 19.8% in Caucasia and Central Asia, respectively) was much higher than in other regions.

Fifteen alleles were found at the *Glu-D1* locus. Among them, the *Glu-D1a* allele occurred in 81.5% of the accessions. This allele was the major allele in all regions except for Caucasia. In Caucasia, the frequency of *Glu-D1a* was only 23.1% while the *Glu-D1d* allele had a frequency of 61.5%. Central Asia showed a tendency similar to Caucasia. In Central Asia, the most frequent allele was *Glu-D1a* as it was in all regions other than Caucasia, but its frequency was much lower (46.9%) than in the other regions. Furthermore, the frequency of *Glu-D1d* was high (23.5%) as it was in Caucasia. At the *Glu-D1* locus, a new allele, *Glu-D1bs*, was discovered. This allele encodes a novel combination of Glu-D1 subunits, 2.1+12. To date, it was thought that the subunit 2.1 was Afghanistan-specific (Lagudah et al. 1987), but in this study, it was found in nine and 19 accessions from Central Asia and Eastern Asia, respectively. As well as *Glu-D1bs*, several of the alleles detected in this study at the *Glu-D1* locus, i.e., *Glu-D1i*, *Glu-D1j*, *Glu-D1l*, *Glu-D1m*, *Glu-D1n*, and *Glu-D1br*, were Asian specific alleles according to previous studies (Gregova et al. 1999, 2006).

In total, 83 allelic combinations of the three *Glu-1* loci were found, among which the most frequent genotype was the combination *Glu-A1c*, *Glu-B1b*, and *Glu-D1a* (Table 3). This genotype occurred in 55.2% of the accessions, and all of the other genotypes had a frequency of less than 10%. The genotype *Glu-A1c*, *Glu-B1b*, and *Glu-D1a* occurred in at least 10% of the accessions in all regions, but its frequency differed greatly among the regions. The

Table 2 Allelic variation at the *Glu-1* loci in Asian wheat accessions

Locus	Allele	Subunit	Western Asia[a]		Caucasia		Central Asia		Afghanistan[a]		Southern Asia[a]		Eastern Asia		Total	
			n[b]	P[c]	n	P	n	P	n	P	n	P	n	P	n	P
Glu-A1	*a*	1	6	8.8	2	7.7	14	17.3	35	8.5	8	8.8	30	6.5	95	8.3
	b	2*	7	10.3	8	30.8	14	17.3	69	16.8	31	34.4	68	14.8	197	17.3
	c	null	55	80.9	16	61.5	53	65.4	306	74.6	54	60.0	363	78.7	847	74.4
Glu-B1	*a*	7	0	0.0	9	34.6	16	19.8	5	1.2	8	8.9	34	7.3	72	6.3
	b	7+8	50	73.5	11	42.3	49	60.5	358	87.3	35	38.9	367	79.6	870	76.5
	c	7+9	2	2.9	3	11.5	7	8.6	13	3.2	3	3.3	11	2.4	39	3.4
	d	6+8	1	1.5	0	0.0	2	2.5	10	2.4	0	0.0	18	3.9	31	2.7
	e	20	7	10.3	3	11.5	4	4.9	3	0.7	3	3.3	7	1.9	27	2.4
	f	13+16	5	7.4	0	0.0	0	0.0	0	0.0	1	1.1	1	0.2	7	0.6
	i	17+18	3	4.4	0	0.0	0	0.0	5	1.2	42	46.7	14	3.0	64	5.6
	aj	8	0	0.0	0	0.0	3	3.7	15	3.7	1	1.1	9	1.9	28	2.4
	ah	null	0	0.0	0	0.0	0	0.0	1	0.2	0	0.0	0	0.0	1	0.1
Glu-D1	*a*	2+12	58	85.3	6	23.1	38	46.9	351	85.6	72	77.4	403	87.4	928	81.5
	b	3+12	0	0.0	0	0.0	0	0.0	2	0.5	0	0.0	0	0.0	2	0.2
	c	4+12	3	4.4	0	0.0	0	0.0	0	0.0	3	3.2	0	0.0	6	0.5
	d	5+10	3	4.4	16	61.5	19	23.5	11	2.7	8	8.6	29	6.3	86	7.5
	e	2+10	2	2.9	2	7.7	1	1.2	3	0.7	2	2.2	0	0.0	10	0.9
	h	5+12	0	0.0	0	0.0	3	3.7	3	0.7	3	3.2	1	0.2	10	0.9
	i	null	0	0.0	0	0.0	0	0.0	0	0.0	3	3.2	0	0.0	3	0.3
	J	2+12*	0	0.0	0	0.0	5	6.2	6	1.5	0	0.0	0	0.0	11	1.0
	k	2	0	0.0	0	0.0	1	1.2	2	0.5	0	0.0	1	0.2	4	0.3
	l	12	2	2.9	1	3.8	0	0.0	6	1.5	1	1.1	8	1.7	18	1.6
	n	2.1+10	0	0.0	0	0.0	8	9.9	0	0.0	0	0.0	17	3.7	25	2.2
	m	10	0	0.0	1	3.8	5	6.2	25	6.1	1	1.1	0	0.0	32	2.8
	br	2.8+12	0	0.0	0	0.0	0	0.0	1	0.3	0	0.0	0	0.0	1	0.1
	bs	2.1+12	0	0.0	0	0.0	1	1.2	0	0.0	0	0.0	2	0.4	3	0.3

[a] All the data for Afghanistan and some of the data for Western Asia and Southern Asia were derived from Terasawa et al. (2009)

[b] Number of accessions possessing the allele

[c] Frequency shown as a percentage

frequency of the genotype was more than 60% in Western Asia, Afghanistan, and Eastern Asia, whereas it was less than 30% in Caucasia, Central Asia, and Southern Asia. In Caucasia and Southern Asia in particular, the most frequent genotype was *Glu-A1b*, *Glu-B1a*, and *Glu-D1d* and *Glu-A1c*, *Glu-B1i*, and *Glu-D1a*, respectively.

Genetic diversity of HMW glutenin subunit composition

The PIC values calculated based on the allelic frequencies at each of the loci are shown in Table 4. The level of diversity was similar among the loci. The PIC value for *Glu-D1* was slightly lower than those for the other two loci in spite of the fact that it had the largest number of alleles due to the high frequency of the major allele. When we compared the level of diversity among the regions, Caucasia, Central Asia, and Southern Asia showed a higher level of variation than Western Asia, Afghanistan, and Eastern Asia at all loci. Based on the frequencies of allelic combination of the three *Glu-1* loci, the highest level of diversity was detected in Caucasia and Central Asia (PIC = 0.90), followed by Southern Asia (PIC = 0.89). The PIC values for the other three regions were approximately 0.6.

Table 3 The frequencies of major genotypes of HMW glutenin subunits in the Asian endemic wheats examined in this study

Genotype	Subunit	Western Asia	Afghanistan	Caucasia	Central Asia	Southern Asia	Eastern Asia	Total
Glu-A1c, *Glu-B1b*, *Glu-D1a*	null, 7+8, 2+12	64.7	65.9	11.5	25.9	14.0	60.3	55.2
Glu-A1b, *Glu-B1b*, *Glu-D1a*	2*, 7+8, 2+12	4.4	12.0	0.0	4.9	6.5	9.1	9.1
Glu-A1c, *Glu-B1a*, *Glu-D1z*	null, 7, 2+12	0.0	0.2	3.8	0.0	3.2	4.3	2.2
Glu-A1c, *Glu-B1d*, *Glu-D1a*	null, 6+8, 2+12	1.5	2.0	0.0	1.2	0.0	3.3	2.2
Glu-A1c, *Glu-B1i*, *Glu-D1a*	null, 17+18, 2+12	0.0	0.0	0.0	0.0	22.6	0.4	2.0
Glu-A1b, *Glu-B1a*, *Glu-D1d*	2*, 7, 5+10	0.0	0.5	23.1	3.7	2.2	1.1	1.6
Glu-A1c, *Glu-B1b*, *Glu-D1d*	null, 7+8, 5+10	0.0	0.2	15.4	1.2	0.0	2.4	1.5
Glu-A1c, *Glu-B1b*, *Glu-D1n*	null. 7+8, 2.1+10	0.0	0.0	0.0	9.9	0.0	1.7	1.4
Glu-A1b, *Glu-B1i*, *Glu-D1a*	2*, 17+18, 2+12	0.0	0.0	0.0	0.0	14.0	0.2	1.2
Glu-A1c, *Glu-B1e*, *Glu-D1a*	null, 20, 2+12	4.4	0.2	3.8	0.0	2.2	0.9	1.0

The genotypes with an overall frequency of more than 2% and the most and second most frequent genotypes in a region are shown. Each frequency is shown as a percentage

Table 4 The PIC values estimated based on the allelic frequency of each *Glu-1* loci

Region	*Glu-A1*	*Glu-B1*	*Glu-D1*	Combination[a]
Western Asia	0.33	0.44	0.27	0.58
Caucasia	0.52	0.68	0.61	0.90
Central Asia	0.51	0.58	0.71	0.90
Afghanistan	0.40	0.22	0.27	0.57
Southern Asia	0.51	0.62	0.35	0.89
Eastern Asia	0.35	0.36	0.23	0.62
Total	0.41	0.41	0.32	0.68

[a] The PIC values were calculateted based on the frequencies of allelic combination of the three *Glu-1* loci

Discussion

In this study, we revealed the genetic variation of HMW glutenin subunit composition in 1,139 wheat accessions endemic to Asia. The genetic constitutions of the wheats endemic to Western Asia, Afghanistan, and Eastern Asia were similar to one another, while those from Caucasia and Central Asia showed different features from these three regions, and those from Southern Asia had a unique genetic constitution.

As reported in our previous paper about Afghan wheat (Terasawa et al. 2009) and that of Lagudah et al. (1987), the genotype *Glu-A1c*, *Glu-B1b*, and *Glu-D1a* was the most common in Western and Eastern Asia. Other studies have indicated that this genotype was also the most common in Japan and the Xinjian District of China (Nakamura 2000; Cong et al. 2005). These results indicated that the genotype *Glu-A1c*, *Glu-B1b*, and *Glu-D1a* is the dominant genotype in the central belt of Asia stretching from Mesopotamia to the Far East via Afghanistan. The wide-ranging high frequency distribution of the genotype is considered to represent the historically flourishing trade and active cultural exchange between Western and Eastern Asia; i.e., "the Silk Road". The genotype *Glu-A1c*, *Glu-B1b*, and *Glu-D1a* encodes subunits null, 7+8, and 2+12. This subunit composition is considered to be unsuitable for making bread. This is probably because of the preference for non-fermented flour foods such as flat bread and noodles in Asia. Indeed, this genotype is not common in European endemic wheat or modern cultivars (Lee et al. 1995; Gregova et al. 1999, 2006; Wrigley et al. online). The genotype was not always the most common but was found at various frequencies in all regions of Asia. Consequently, we conclude that the allelic combination of *Glu-A1c*, *Glu-B1b*, and *Glu-D1a* is the typical genotype that characterizes Asian endemic wheat.

In Southern Asia, the most frequent genotype was *Glu-A1c*, *Glu-B1i*, and *Glu-D1a*. This genotype can be regarded as a modified version of the typical Asian genotype that was generated by replacing the *Glu-B1b* with *Glu-B1i*. In this study, the *Glu-B1i* allele was quite rare in regions other than Southern Asia. Similarly, this allele was seldom discovered in European endemic wheat (Gregova et al. 1999, 2006; Juhász et al. 2003). Hence, we consider that

the *Glu-B1i* allele was generated in Southern Asia. The *Glu-B1i* allele encodes subunit 17+18, which produces stronger dough than the *Glu-B1b* allele (Payne and Lawrence 1983; Mondal et al. 2008). Since stronger dough is preferable even for semi-fermented flat bread such as naan, which is commonly eaten in Southern Asia, the *Glu-B1i* allele might be subject to positive selection in this region.

The accessions from Caucasia and Central Asia had a unique genetic constitution that was different from those of the other regions. The high frequency of the *Glu-D1d* allele was the most striking feature of the wheats from these regions, although high proportions of the *Glu-A1b* and *Glu-B1a* alleles were also characteristic of these regions. The subunit 5+10, which is encoded by the *Glu-D1d* allele, is known to improve bread-making quality and so has been intensively introduced into a number of modern cultivars (Lee et al. 1995; Wrigley et al. online). Although both Caucasia and Central Asia used to belong to the USSR, the genetic constitutions of the *Glu-1* loci in these regions were different from those of the former USSR modern varieties reported by Morgunov et al. (1990). Therefore, the high frequency of *Glu-D1d* in these regions did not result from modern breeding. The *Glu-D1d* allele is also frequent in European endemic wheat (Gregova et al. 1999, 2006; Juhász et al. 2003). However, considering the fact that the highest frequency of *Glu-D1d* was detected in Caucasia and that bread wheat originated near to Caucasia (Dvorak et al. 1998), it is possible that the *Glu-D1d* allele was originally generated in Caucasia and spread to Europe and other corners of the world gradually.

The level of genetic diversity was relatively low in the central belt of Asia (Western Asia, Afghanistan, and Eastern Asia) and high in the marginal regions (Caucasia, Central Asia, and South Asia). This result arose from differences in the frequency of the typical Asian genotype. The levels of genetic diversity in the northern regions (Caucasia and Central Asia) and Southern Asia were similar to each other, but their genetic constitutions were quite different, as described above. Since modern breeding had little influence on our samples, the genetic differentiation among the regions was probably produced by the limited exchange of genetic resources and the accumulation of specific alleles in the marginal regions. Even though their frequencies were low, the presence of Asian specific alleles, especially at the *Glu-D1* locus, contributes to the unique genetic properties of Asian endemic wheat. Contrary to the findings of the present study, European endemic wheat has higher numbers of alleles at the *Glu-B1* locus than at the *Glu-D1* locus (Gregova et al. 1999, 2006). This might be due to differences in the requirement for the *Glu-D1d* allele encoding the subunit 5+10, which is preferred for making bread, between Asia and Europe.

In conclusion, this study comprehensively clarified the genetic properties of HMW glutenin subunit composition in Asian endemic wheats. As far as we know, this is the first study to investigate such a large number of accessions of Asian endemic wheats from such a wide-ranging distribution. The results obtained in this study will help to elucidate how wheat spread throughout Asia and aid future breeding.

Acknowledgments This work was supported by the National BioResources Project (NBRP) KOMUGI and a Grant-in-Aid for Scientific Research (A) (10201047). We express our sincere gratitude to Drs. Tatsuya Ikeda, Hisashi Tsujimoto, Motoko Takaoka, and Toshinori Abe for their technical support and critical discussion. Finally, we thank all the members of the Laboratory of Evolutionary Genetics of the Kihara Institute for Biological Research for their kind help and continuous encouragement throughout this study.

References

Cong H, Takata K, Zong YF, Ikeda TM, Yanaka M, Nagamine T, Fujimaki H (2005) Novel high-molecular-weight glutenin subunits at the *Glu-D1* locus in wheat landraces from the Xinjiang District of China and relationship with winter habit. Breed Sci 55:459–463

Dvorak J, Luo MC, Yang ZL, Zhang HB (1998) The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet 97:657–670

Fang J, Liu Y, Luo J, Wang Y, Shewry PR, He G (2009) Allelic variation and genetic diversity of high molecular weight glutenin subunit in Chinese endemic wheats (*Triticum aestivum* L.). Euphytica 166:177–182

Gregova E, Hermuth J, Kraic J, Dotlacil L (1999) Protein heterogeneity in European wheat landraces and obsolete cultivars. Additional information. Genet Resour Crop Evol 46:521–528

Gregova E, Hermuth J, Kraic J, Dotlacil L (2006) Protein heterogeneity in European wheat landraces and obsolete cultivars. Additional information II. Genet Resour Crop Evol 53:867–871

Juhász A, Larroque OR, Tamás L, Hsam S, Zeller FJ, Békés F, Bedo Z (2003) Bánikúti 1201—an old Hungarian wheat variety with special storage protein composition. Theor Appl Genet 107:697–704

Lagudah ES, Flood RG, Halloran GM (1987) Variation in high molecular weight glutenin subunits in landraces of hexaploid wheat from Afghanistan. Euphytica 36:3–9

Lee CW, Ito S, Watanabe M, Hoshino T (1995) Composition of high molecular weight–glutenin subunit in wheat cultivers. Misc Publ Tohoku Natl Agric Exp Stn 17:33–40

Mondal S, Tilley M, Alviola JN, Waniska RD, Bean SR, Glover KD, Hays DB (2008) Use of near-isogenic wheat lines to determine the glutenin composition and functionality requirements for flour tortillas. J Agric Food Chem 56:179–184

Morgunov AI, Rogers WJ, Sayers EJ, Metakovsky EV (1990) The high-molecular-weight glutenin subunit composition of Soviet wheat varieties. Euphytica 51:41–52

Nakamura H (2000) Allelic variation at high-molecular-weight glutenin subunit loci, *GluA1*, *GluB1* and *GluD1*, in Japanese and Chinese hexaploid wheats. Euphytica 112:187–193

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Payne PI, Lawrence GJ (1983) Catalogue of alleles for the complex gene loci, *Glu-A1*, *Glu-B1*, and *Glu-D1* which code for high molecular-weight subunits of glutenin in hexaploid wheat. Cereal Res Commun 11:29–35

Payne PI, Law CN, Mudd EE (1980) Control by homoeologous group 1 chromosomes of the high-molecular-weight subunits of glutenin, a major protein of wheat endosperm. Theor Appl Genet 58:113–120

Payne PI, Holt LM, Worland AJ, Law CN (1982) Structural and genetical studies on the high-molecular-weight subunits of glutenin. 3. Telocentric mapping of subunit genes on the long arms of the homoeologous group 1 chromosomes. Theor Appl Genet 63:129–138

Shewry PR, Halford NG, Tatham AS (1992) High molecular weight subunits of wheat glutenin. J Cereal Sci 15:105–120

Tahir M, Turchetta T, Anwar R, Lafiandra D (1996) Assessment of genetic variability in hexaploid wheat landraces of Pakistan based on polymorphism for HMW glutenin subunits. Genet Resour Crop Evol 43:211–220

Terasawa Y, Kawahara T, Sasakuma T, Sasanuma T (2009) Evaluation of the genetic diversity of an Afghan wheat collection based on morphological variation, HMW glutenin subunit polymorphisms, and AFLP. Breed Sci 59:361–371

Vavilov NI (1926) Studies on the origins of cultivated plants. Institute of Applied Botany and Plant Breeding, Leningrad

Wrigley CW, Bekes F, Cavanagh CR, Bushuk W (online) The gluten composition of wheat varieties and genotypes. http://www.aaccnet.org/GRAINBIN/gluten_gliadin.asp

Zhang XU, Pang BS, You GX, Wang LF, Jia JZ, Dong YC (2002) Allelic variation and genetic diversity at *Glu-D1* loci in Chinese wheat (*Triticum aestivum* L.) germplasms. Sci Agric Sin 35:1302–1310 (in Chinese with English summary)

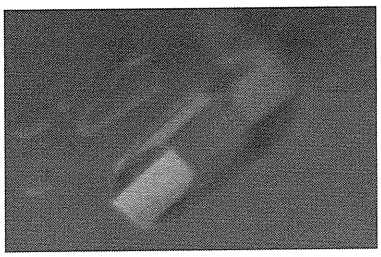
Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Takuya Yoshizawa,[a] Toshiyuki Shimizu,[b] Hisashi Hirano,[a] Mamoru Sato[a] and Hiroshi Hashimoto[a]*

[a]Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan, and [b]Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Correspondence e-mail: hash@tsurumi.yokohama-cu.ac.jp

Received 7 April 2011
Accepted 26 May 2011



Purification, crystallization and X-ray diffraction study of extracellular dermal glycoprotein from carrot and the inhibition complex that it forms with an endo-β-glucanase from *Aspergillus aculeatus*

Extracellular dermal glycoprotein (EDGP) may play an important role in the plant defence system of the carrot (*Daucus carota*) as it has inhibitory activity against endo-β-glucanase produced by invading pathogens. Here, EDGP and the inhibition complex that it forms with FI-CMCase, a carboxyl methyl cellulase from *Aspergillus aculeatus*, were successfully crystallized. The hexagonal crystal of EDGP belonged to space group $P6_2$, with unit-cell parameters $a = b = 130.4$, $c = 44.5$ Å, $\gamma = 120°$. The monoclinic crystal of the complex of EDGP with FI-CMCase belonged to space group $C2$, with unit-cell parameters $a = 169.5$, $b = 143.0$, $c = 63.0$ Å, $\beta = 110.9°$.

1. Introduction

Plant cell walls are composed of various polysaccharides such as cellulose, hemicellulose and pectin. Cellulose microfibrils are linked *via* hemicellulose. This cellulose–hemicellulose network provides tensile strength and acts as a physical barrier against invading pathogens. To penetrate and use plant cell walls nutritionally, pathogens secrete cell-wall-degrading enzymes. These enzymes, which include endoglucanases, xylanases and polygalacturonases, are classified into glycoside hydrolase (GH) families in the CAZY database (Henrissat, 1991). In response to pathogenic attack, plants produce proteinous inhibitors against these cell-wall-degrading enzymes (Juge, 2006; Lagaert *et al.*, 2009). Extracellular dermal glycoprotein (EDGP) is one such proteinous inhibitor against these degrading enzymes. EDGP shows inhibitory activity towards the xyloglucan-specific endo-β-1,4-glucanase (XEG) from *Aspergillus aculeatus* (Shang *et al.*, 2005), which belongs to the GH12 family and specifically cleaves xyloglucan, which is a major hemicellulose found in dicots (Pauly *et al.*, 1999). Xyloglucan consists of a β-linked glucose backbone substituted with xylose side chains. The degradation of xyloglucan causes great damage to dicotyledonous plants. Consequently, the inhibition of XEG by EDGP is an important component of the defence system of the carrot.

Homologous proteins to EDGP are widely present in plants and several of these homologous proteins have been characterized. Tomato (*Lycopersicon esculentum*) XEGIP (xyloglucan-specific endo-β-1,4-glucanase inhibitor protein) inhibits XEG by forming an associated 1:1 complex (Qin *et al.*, 2003). Necturin IV (NEC IV) from tobacco (*Nicotiana langsdorffii* × *N. sanderae* var. LxS8) also inhibits XEG (Naqvi *et al.*, 2005; Sansen *et al.*, 2004). In contrast, the homologous protein from wheat (*Triticum aestivum* xylanase inhibitor; TAXI) inhibits GH11 xylanases (Gebruers *et al.*, 2004; Sansen *et al.*, 2004). Interestingly, the soybean (*Glycine max* L. Merrill cv. Miyagishirome) homologue basic 7S globulin (Bg7S) lacks inhibitory activity for both GH11 and GH12 enzymes (Yoshizawa *et al.*, 2011). EDGP shares 61, 62, 21 and 37% amino-acid identity with tomato XEGIP, tobacco NEC IV, wheat TAXI-IA and soybean Bg7S, respectively. Structural analyses of TAXI and Bg7S have already been reported (Sansen *et al.*, 2004; Yoshizawa *et al.*, 2011). However, structures have not been determined for EDGP and homologous proteins that inhibit GH12 enzymes.

Expression of EDGP is induced by both biotic and abiotic stress (Satoh *et al.*, 1986). EDGP consists of 413 amino-acid residues and is

subject to post-translational modifications; the processed protein contains N-terminal pyroglutamic acid, six disulfide-bond pairs and four N-linked glycosylation sites (Shang *et al.*, 2004). Deglycosylation of EDGP causes a complete loss of its ability to inhibit XEG (Shang *et al.*, 2005). To clarify the inhibition mechanism of EDGP towards GH12 endo-β-glucanase, we used X-ray crystallography to determine the structure of both EDGP and EDGP in complex with a GH12 enzyme. In the present paper, we describe the purification and crystallization of EDGP and EDGP in complex with FI-CMCase, a GH12 endo-β-glucanase from *A. aculeatus* (Kanda *et al.*, 1976), and also detail the initial diffraction analysis of these crystals.

2. Materials and results

2.1. Preparations of EDGP and FI-CMCase for crystallographic study

EDGP was purified from carrot callus culture medium. Carrot callus was grown for 2–3 weeks at 298 K in Murashige–Skoog (MS) medium containing 1 mg l^{-1} 2,4-dichlorophenoxyacetic acid. The carrot callus culture suspension was filtered using Miracloth (Merck KGaA) and clarified by centrifugation for 30 min at 277 K (43 667*g*). An equal volume of 50 m*M* sodium acetate pH 4.6 was added to the supernatant and it was applied onto a HiTrap SP HP column (GE Healthcare) equilibrated with 50 m*M* sodium acetate pH 4.6. The bound proteins were eluted with a linear gradient from 0 to 400 m*M* NaCl. Fractions containing EDGP were concentrated to 20 mg ml^{-1} in 20 m*M* potassium phosphate pH 7.4 using Amicon Ultra 30 kDa molecular-weight cutoff filter units (Millipore). The purity of EDGP was confirmed by SDS–PAGE with Coomassie Brilliant Blue (CBB) staining (Fig. 1, lane 1). About 1 mg purified EDGP was obtained from 400 ml culture.

The cDNAs encoding FI-CMCase were obtained by PCR-based gene synthesis and inserted into pGEX6P-I vector (GE Healthcare) at the *Bam*HI–*Xho*I site. The plasmid encodes FI-CMCase with a GST tag at the N-terminus. The expression vector was introduced into *Escherichia coli* strain BL21. The cells were grown at 310 K to an optical density of 0.6 at 660 nm in LB medium containing 50 μg ml^{-1} ampicillin. After the addition of 1 m*M* isopropyl β-D-1-thiogalacto-pyranoside (IPTG), cells were grown for 6 h at 298 K. The cells were then harvested, suspended in lysis buffer (50 m*M* HEPES–NaOH pH 7.4, 250 m*M* NaCl, 5 m*M* EDTA and 0.5 m*M* PMSF) and disrupted by sonication. After centrifugation, the supernatant was applied onto glutathione Sepharose 4B (GS4B) resin (GE Healthcare). The bound proteins were eluted with 50 m*M* Tris–HCl pH 9.0, 200 m*M* NaCl and 30 m*M* reduced glutathione. Fractions containing the GST-fused FI-CMCase were dialyzed with 50 m*M* Tris–HCl pH 9.0 and 200 m*M* NaCl. The N-terminal GST tag was cleaved by HRV3C protease at 277 K for 15 h and the solution was then applied onto GS4B resin to remove GST. The flowthrough fractions containing FI-CMCase were applied onto a HiLoad Superdex 75 column (GE Healthcare) equilibrated with 5 m*M* Tris–HCl pH 9.0 and 100 m*M* NaCl. Fractions containing FI-CMCase were concentrated to 20 mg ml^{-1} using Amicon Ultra 10 kDa molecular-weight cutoff filter units (Millipore). The purity of FI-CMCase was confirmed by SDS–PAGE with CBB stain (Fig. 1, lane 2). The EDGP–FI-CMCase complex was prepared by mixing EDGP and FI-CMCase in an equimolar ratio (13 mg ml^{-1} EDGP and 7 mg ml^{-1} FI-CMCase in 13 m*M* potassium phosphate pH 7.4, 2 m*M* Tris–HCl pH 9.0 and 35 m*M* NaCl).

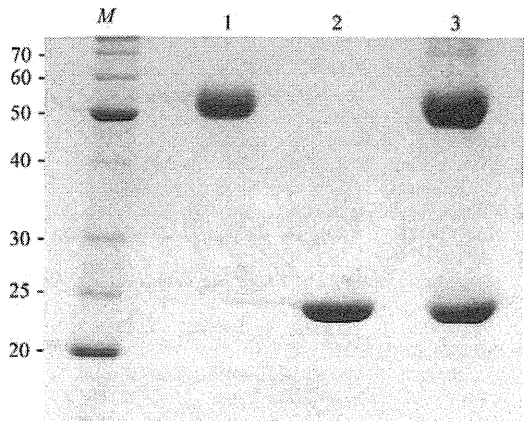


Figure 1
SDS–PAGE analysis of purified EDGP (lane 1), purified FI-CMCase (lane 2) and a crystal of the EDGP–FI-CMCase complex (lane 3). Lane *M* contains a molecular-size marker (labelled in kDa).

2.2. Crystallization and initial crystallographic study of EDGP and the EDGP–FI-CMCase complex

Screening of crystallization conditions for EDGP and the EDGP–FI-CMCase complex was performed by the sitting-drop vapour-diffusion method using a Hydra II Plus One (Matrix) and more than

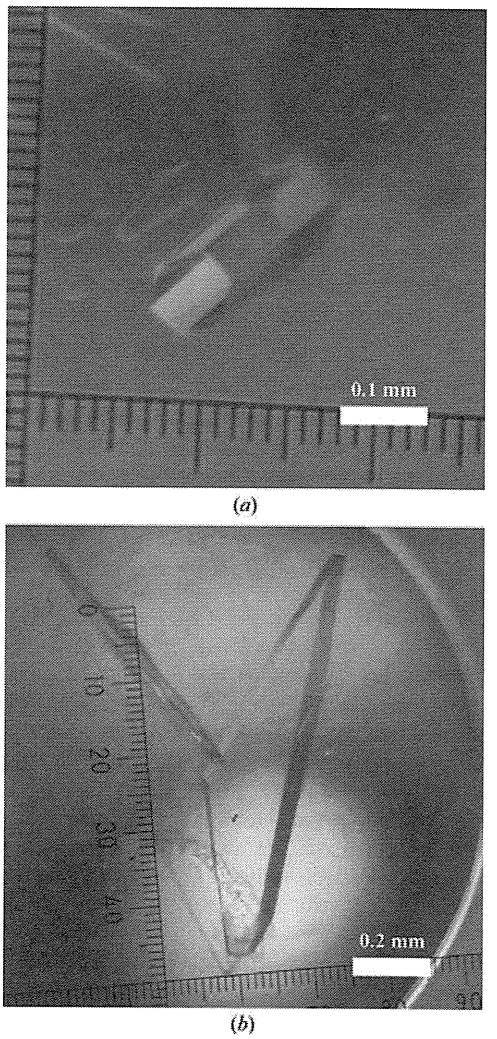


Figure 2
Crystals of EDGP (*a*) and the EDGP–FI-CMCase complex (*b*).

Table 1
Data-collection statistics for crystals of EDGP and the EDGP–FI-CMCase complex.

Values in parentheses are for the highest resolution shell.

		EDGP (Pt derivative)				
	EDGP (Native)	Peak	Edge	High remote	Low remote	EDGP–FI-CMCase
Wavelength (Å)	0.9000	1.07244	1.07171	1.05420	1.07710	1.0000
Resolution range (Å)	50.0–1.00 (1.04–1.00)	50.0–1.80 (1.86–1.80)	50.0–1.80 (1.86–1.80)	50.0–1.80 (1.86–1.80)	50.0–1.80 (1.86–1.80)	50.0–2.50 (2.59–2.50)
Measured reflections	843817 (41436)	439831 (40733)	441642 (37085)	427386 (25357)	409988 (15200)	185159 (18464)
Unique reflections	223206 (19344)	40397 (4033)	40518 (4031)	40545 (3962)	40315 (3619)	48804 (4859)
Multiplicity	3.8 (2.2)	10.9 (10.1)	10.9 (9.2)	10.5 (6.4)	10.2 (4.2)	3.8 (3.8)
Completeness (%)	96.2 (83.5)	100.0 (100.0)	100.0 (100.0)	99.9 (98.8)	98.9 (89.4)	99.9 (100.0)
Mean $I/\sigma(I)$	10.5 (1.5)	16.5 (6.4)	15.5 (5.0)	14.1 (3.1)	13.2 (1.9)	14.4 (5.5)
R_{merge}† (%)	6.0 (37.0)	8.3 (30.6)	8.1 (35.4)	8.6 (44.0)	8.6 (52.6)	7.1 (30.2)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

480 different reservoir solutions from commercially available screening kits (Hampton Research, Molecular Dimensions and Qiagen) at 293 K. Crystals of EDGP were obtained using the following reservoir solutions: No. 10 from Crystal Screen (0.2 *M* ammonium acetate, 0.1 *M* sodium acetate pH 4.6 and 30% PEG 4000), No. 82 from the JCSG+ suite (0.15 *M* potassium bromide and 30% PEG MME 2000) and No. 94 from the JCSG+ suite (0.2 *M* ammonium acetate, 0.1 *M* Bis-Tris pH 5.5 and 25% PEG 3350). The buffer, the pH and the salt and precipitant concentrations of these crystallization conditions were optimized using the hanging-drop vapour-diffusion method. Hexagonal crystals of EDGP suitable for diffraction studies were obtained using a reservoir solution consisting of 0.2 *M* ammonium acetate, 0.1 *M* sodium acetate pH 4.6 and 30% PEG MME 2000 after several days (Fig. 2*a*). Prior to the X-ray experiment, the EDGP crystal was transferred into a cryoprotectant solution consisting of 0.2 *M* ammonium acetate, 0.1 *M* sodium acetate pH 4.6 and 40% PEG MME 2000 with a nylon loop and then cooled in an N_2-gas stream at 100 K. X-ray diffraction data for the EDGP crystal were collected using a Quantum 315 CCD detector (Area Detector Systems Corp.) on beamline BL-41XU at SPring-8 (Sayo, Japan). Diffraction data were integrated, scaled and averaged with the *HKL*-2000 program suite (Otwinowski & Minor, 1997). The EDGP crystal belonged to space group $P6_2$, with unit-cell parameters $a = b = 130.4$, $c = 44.5$ Å, $\gamma = 120°$. Data-collection statistics are summarized in Table 1. The asymmetric unit of the EDGP crystal is estimated to contain a single molecule ($V_M = 2.43$ Å^3 Da^{-1}). EDGP contains 12 cysteines and six disulfide-bond pairs. Therefore, a heavy-atom derivative was prepared by soaking the native crystal in reservoir solution containing 10 m*M* K_2PtCl_4 for 15 h. X-ray diffraction data for the platinum-derivative crystal were collected using a Quantum 270 CCD detector (Area Detector Systems Corp.) on beamline BL-17A at Photon Factory (PF). The wavelengths for MAD measurements were selected based on XAFS measurements. Crystals of the EDGP–FI-CMCase complex were obtained by initial screening using the following reservoir solutions: No. 89 from the PACT suite (0.2 *M* sodium nitrate, 0.1 *M* Bis-Tris propane pH 8.5 and 20% PEG 3350) and No. 9 from the Protein Complex suite (0.2 *M* sodium chloride, 0.1 *M* MES pH 6.0 and 20% PEG MME 2000). After these crystallization conditions had been optimized, a monoclinic crystal of the EDGP–FI-CMCase complex suitable for diffraction study was obtained using a reservoir solution consisting of 0.4 *M* sodium chloride, 0.1 *M* MES pH 6.0, 22% PEG MME 2000 and 5% glycerol (Fig. 2*b*) after 10 d. The crystal was confirmed to contain EDGP and FI-CMC by SDS–PAGE using CBB stain (Fig. 1, lane 3). For cryoprotection, the EDGP–FI-CMCase complex crystal was transferred into a cryoprotectant solution consisting of 0.4 *M* sodium chloride, 0.1 *M* MES pH 6.0, 22% PEG MME 2000 and 20% glycerol. X-ray diffraction data for the EDGP–FI-CMCase complex crystal were collected using a Quantum 270 CCD detector (Area Detector Systems Corp.) on the NE-3A beamline at Photon Factory (PF). The crystal belonged to space group *C*2, with unit-cell parameters $a = 169.5$, $b = 143.0$, $c = 63.0$ Å, $\beta = 110.9°$. The asymmetric unit was estimated to contain two EDGP–FI-CMCase complexes ($V_M = 2.58$ Å^3 Da^{-1}). Structure determination of EDGP and the EDGP–FI-CMCase complex is now in progress.

We acknowledge the kind support of the beamline staff at SPring-8 and PF. This work was supported by grants from KAKENHI, the National Project on Protein Structural and Functional Analyses (Protein 3000 Project) and the Targeted Proteins Research Program to MS, TS and HH from MEXT. We thank Dr H. Shiota, Yokohama City University for supplying carrot callus and Dr C. Everroad, Riken Yokohama Institute for English corrections.

References

Gebruers, K., Brijs, K., Courtin, C. M., Fierens, K., Goesaert, H., Rabijns, A., Raedschelders, G., Robben, J., Sansen, S., Sørensen, J. F., Van Campenhout, S. & Delcour, J. A. (2004). *Biochim. Biophys. Acta*, **1696**, 213–221.
Henrissat, B. (1991). *Biochem. J.* **280**, 309–316.
Juge, N. (2006). *Trends Plant Sci.* **11**, 359–367.
Kanda, T., Wakabayashi, K. & Nisizawa, K. (1976). *J. Biochem.* **79**, 997–1005.
Lagaert, S., Beliën, T. & Volckaert, G. (2009). *Semin. Cell Dev. Biol.* **20**, 1064–1073.
Naqvi, S. M., Harper, A., Carter, C., Ren, G., Guirgis, A., York, W. S. & Thornburg, R. W. (2005). *Plant Physiol.* **139**, 1389–1400.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pauly, M., Andersen, L. N., Kauppinen, S., Kofod, L. V., York, W. S., Albersheim, P. & Darvill, A. (1999). *Glycobiology*, **9**, 93–100.
Qin, Q., Bergmann, C. W., Rose, J. K., Saladie, M., Kolli, V. S., Albersheim, P., Darvill, A. G. & York, W. S. (2003). *Plant J.* **34**, 327–338.
Sansen, S., De Ranter, C. J., Gebruers, K., Brijs, K., Courtin, C. M., Delcour, J. A. & Rabijns, A. (2004). *J. Biol. Chem.* **279**, 36022–36028.
Satoh, S., Kamada, H., Harada, H. & Fujii, T. (1986). *Plant Physiol.* **81**, 931–933.
Shang, C., Sassa, H. & Hirano, H. (2005). *Biochem. Biophys. Res. Commun.* **328**, 144–149.
Shang, C., Shibahara, T., Hanada, K., Iwafune, Y. & Hirano, H. (2004). *Biochemistry*, **43**, 6281–6292.
Yoshizawa, T., Shimizu, T., Yamabe, M., Taichi, M., Nishiuchi, Y., Shichijo, N., Unzai, S., Hirano, H., Sato, M. & Hashimoto, H. (2011). *FEBS J.* **278**, 1944–1954.

the FEBS Journal

Crystal structure of basic 7S globulin, a xyloglucan-specific endo-β-1,4-glucanase inhibitor protein-like protein from soybean lacking inhibitory activity against endo-β-glucanase

Takuya Yoshizawa[1], Toshiyuki Shimizu[2], Mayuki Yamabe[1], Misako Taichi[3,4], Yuji Nishiuchi[3,4], Naoki Shichijo[1], Satoru Unzai[1], Hisashi Hirano[1], Mamoru Sato[1] and Hiroshi Hashimoto[1]

1 Graduate School of Nanobioscience, Yokohama City University, Japan
2 Graduate School of Pharmaceutical Science, The University of Tokyo, Japan
3 SAITO Research Center, Peptide Institute Inc., Osaka, Japan
4 Graduate School of Science, Osaka University, Japan

Keywords
crystal structure; glucanase inhibitor; legume protein; macromolecular assembly; plant defense

Correspondence
H. Hashimoto, Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan
Fax: +81 45 508 7365
Tel: +81 45 508 7227
E-mail: hash@tsurumi.yokohama-cu.ac.jp

(Received 22 February 2011, revised 25 March 2011, accepted 28 March 2011)

doi:10.1111/j.1742-4658.2011.08111.x

β-Linked glucans such as cellulose and xyloglucan are important components of the cell walls of most dicotyledonous plants. These β-linked glucans are constantly exposed to degradation by various endo-β-glucanases from pathogenic bacteria and fungi. To protect the cell wall from degradation by such enzymes, plants secrete proteinaceous endo-β-glucanases inhibitors, such as xyloglucan-specific endo-β-1,4-glucanase inhibitor protein (XEGIP) in tomato. XEGIPs typically inhibit xyloglucanase, a member of the glycoside hydrolase (GH)12 family. XEGIPs are also found in legumes, including soybean and lupin. To date, tomato XEGIP has been well studied, whereas XEGIPs from legumes are less well understood. Here, we determined the crystal structure of basic 7S globulin (Bg7S), a XEGIP from soybean, which represents the first three-dimensional structure of XEGIP. Bg7S formed a tetramer with pseudo-*222* symmetry. Analytical centrifugation and size exclusion chromatography experiments revealed that the assembly of Bg7S in solution depended on pH. The structure of Bg7S was similar to that of a xylanase inhibitor protein from wheat (*Tritinum aestivum* xylanase inhibitor) that inhibits GH11 xylanase. Surprisingly, Bg7S lacked inhibitory activity against not only GH11 but also GH12 enzymes. In addition, we found that XEGIPs from azukibean, yardlongbean and mungbean also had no impact on the activity of either GH12 or GH11 enzymes, indicating that legume XEGIPs generally do not inhibit these enzymes. We reveal the structural basis of why legume XEGIPs lack this inhibitory activity. This study will provide significant clues for understanding the physiological role of Bg7S.

Database
Coordinates and structure factors have been deposited in the Protein Data Bank Japan (PDBj) (http://www.pdbj.org/) under the accession number 3AUP.

Abbreviations
ANXY, *Aspergillus niger* xylanase; ASA, accessible surface area; AUC, analytical ultracentrifugation; BTB, back-to-back; Bg7S, basic 7S globulin; EDGP, extracellular dermal glycoprotein; FTF, face-to-face; GH, glycoside hydrolase; GST, glutathione-*S*-transferase; IL-1, inhibition loop 1; IL-2, inhibition loop 2; PDB, Protein Data Bank; SEC, size exclusion chromatography; TAXI, *Tritinum aestivum* xylanase inhibitor; XEG, xyloglucan-specific endo-β-1,4-glucanase; XEGIP, xyloglucan-specific endo-β-1,4-glucanase inhibitor protein.

Structured digital abstract

- **Bg7S** binds to **Bg7S** by x-ray crystallography (View interaction)
- **Bg7S** binds to **Bg7S** by cosedimentation in solution (View Interaction 1, 2)
- **Bg7S** binds to **Bg7S** by molecular sieving (View Interaction 1, 2)

Introduction

The cell wall of plants is composed of various polysaccharides, such as cellulose and hemicellulose. Cellulose is a major component of the plant cell wall, and cellulose microfibrils are linked via hemicellulose. The network of cellulose–hemicellulose provides tensile strength. In most dicotyledonous plants, hemicellulose comprises xyloglucan, which consists of a cellulosic backbone substituted with side chains. These β-linked glucans, namely cellulose and xyloglucan, are constantly exposed to degradation by various endo-β-glucanases, such as cellulase and xyloglucanase from pathogenic bacteria and fungi. To protect the cell wall from degradation by such enzymes, plants secrete proteinaceous inhibitors against endo-β-glucanases. The first endo-β-glucanase inhibitor protein to be discovered was the so-called xyloglucan-specific endo-β-1,4-glucanase inhibitor protein (XEGIP) [1], a tomato protein that inhibits fungal xyloglucan-specific endo-β-1,4-glucanase (XEG), an enzyme classified as a member of the glycoside hydrolase (GH)12 family in the CAZy database [2] (http://www.cazy.org). Tomato XEGIP is a basic 51-kDa protein, and, as its name suggests, inhibits XEG by forming a tightly associated 1 : 1 complex with an inhibition constant (K_i) of ~ 0.5 nM. XEGIPs have been discovered in various higher plants [3], and some of these proteins have been characterized. For example, carrot XEGIP is termed extracellular dermal glycoprotein (EDGP). It has been shown that EDGP also inhibits XEG from *Aspergillus aculeatus* [4]. Tobacco XEGIP, termed nectarin IV, has been shown to inhibit XEG and does not inhibit GH11 xylanases [5], although the structures of GH12 and GH11 are very similar.

XEGIPs are structurally related to *Tritinum aestivum* xylanase inhibitor (TAXI), a xylanase inhibitor protein isolated from wheat [6], because both XEGIP and TAXI have 12 cysteines in similar positions. These cysteines form six disulfide bonds in the tertiary structure of TAXI [7]. To date, four TAXI isomers have been identified in wheat (TAXI-IA, TAXI-IB, TAXI-IIA, and TAXI-IIB). TAXI inhibits GH11 xylanase, whereas it inhibits neither GH12 nor GH10 xylanase. A structural study has revealed that TAXI-IA adopts a pepsin fold lacking proteolytic activity [7]. The structure of TAXI-IA in complex with *Aspergillus niger* xylanase (ANXY), a GH11 xylanase from *Aspergillus niger*, coupled with functional studies, has revealed that His374 of TAXI-IA plays a significant role in the inhibition of ANXY, where His374 interacts with the catalytic Glu79 and Glu170 of ANXY [7,8]. Furthermore, it has been reported that the hydrophobic interaction of Leu292 of TAXI-IA with Pro294 of TAXI-IIB regulates the strength of inhibition and specificity for GH11 xylanases [9].

XEGIPs are also found in legumes, including lupin and soybean. γ-Conglutin is a XEGIP found in lupin [3]. In soybean, a XEGIP is the basic 7S globulin (Bg7S) [10]. Soybean Bg7S shares 38% and 37% amino acid identity with tomato XEGIP and EDGP, respectively. Bg7S is initially synthesized as a precursor protein with an N-terminal signal peptide. Bg7S is matured by post-translational modifications: cleavage of the N-terminal 24 residues, formation of disulfide bonds, and cleavage between Ser251 and Ser252, where the numbering starts from the first residues of the matured protein. Mature Bg7S consists of 403 amino acids, and has a molecular mass of 43 kDa; it is composed of 27-kDa (α) and 16-kDa (β) chains [10]. Although tomato XEGIP and EDGP are monomeric proteins, Bg7S exists as an oligomeric form [10,11]. Furthermore, it has been reported that Bg7S binds a 4-kDa hormone-like peptide, termed leginsulin, from soybean [11–13]. However, both the structure and function of Bg7S remain unknown. Here, we report the crystal structure of Bg7S from soybean, and functional analysis of Bg7S.

XEGIPs have been discovered in various plants, including potato (Uniprot ID Q7XJE7; sequence identity with Bg7S, 39%), *Arabidopsis* (Q8LF70, 38%), rice (A2Y4I2, 36%), and maize (B6UHL4, 26%). Thus, our structural and functional studies on Bg7S will shed light on XEGIPs which are widely conserved in various plants.

Results and Discussion

Structure of Bg7S from soybean

The crystal structure of soybean Bg7S was determined at 1.9-Å resolution. The asymmetric unit contained four Bg7S protomers (A, B, C and D molecules), and

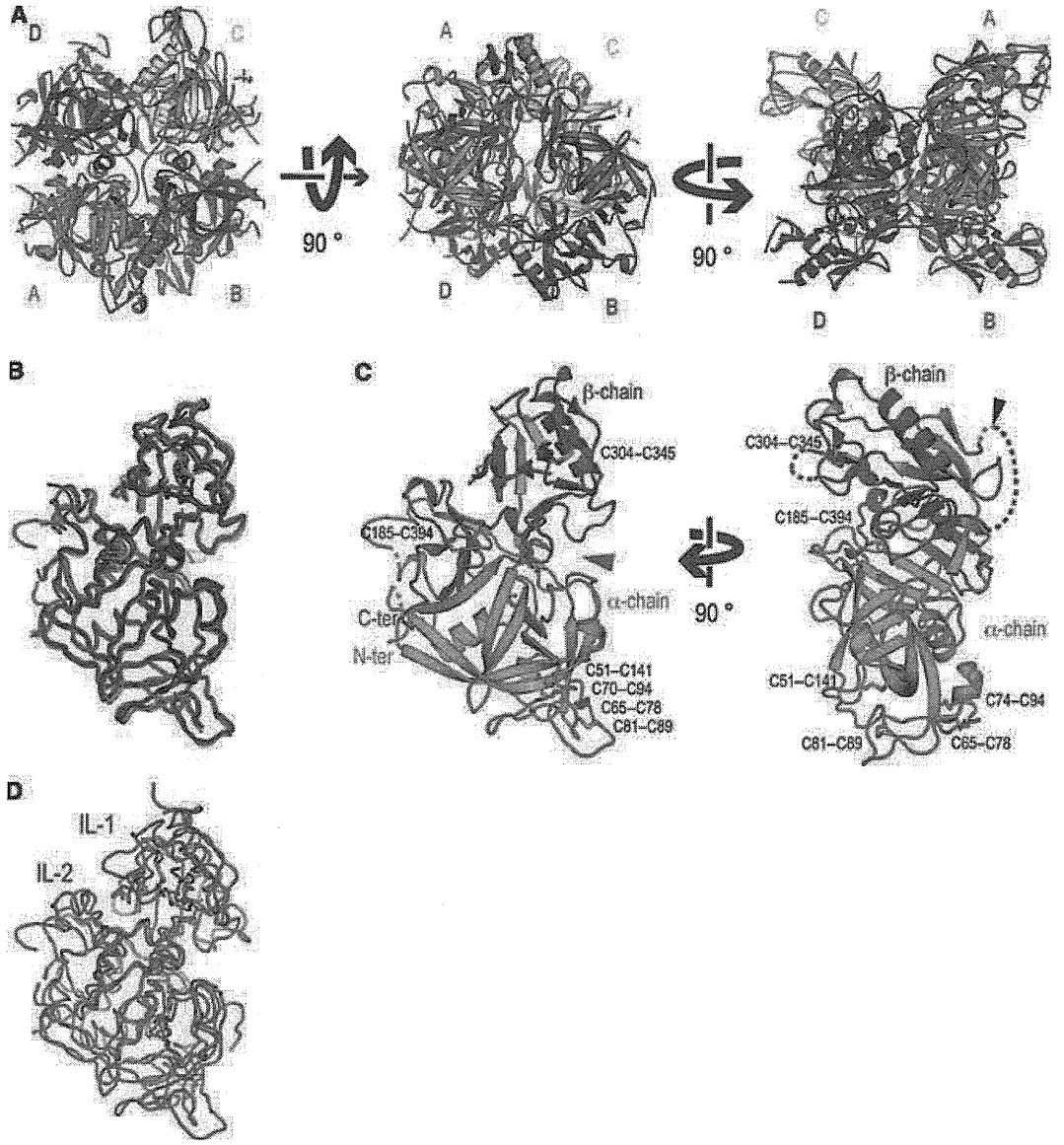


Fig. 1. Structure of Bg7S from soybean. (A) Top and side views of the Bg7S tetramer. A, B, C and D molecules in the asymmetric unit are shown as green, red, yellow and blue ribbon representations, respectively. (B) Superimposed structures of the Bg7S protomers are shown by wire representations. Colors correspond to those in (A). (C) The overall structure of the Bg7S protomer is shown by a ribbon representation. The structure of the A molecule is shown as an example. The N-terminus and C-terminus are labeled. The α-chain and β-chain are shown as green and light blue ribbon representations, respectively. The cysteines involved in the disulfide bonds are shown as stick representations and labeled in black. The disordered regions are shown as dashed lines. The black triangle indicates the post-translational cleavage position. The pseudo-active site of aspartic protease is indicated by the red triangle. (D) Superimposed structures of Bg7S and TAXI-IA (PDB ID 1T6G, chain A) are shown as green and light brown wire representations, respectively. The loops of TAXI-IA involved in interactions with ANXY are labeled IL-1 and IL-2.

they formed a tetramer with pseudo-*222* symmetry (Fig. 1A). The N-terminal moieties of the β-chains of the C and D molecules protrude into the AB dimer (Fig. 1A), whereas the corresponding regions of the A and B molecules are disordered. We have obtained a Bg7S crystal with different cell dimensions [14]: Bg7S also forms a tetramer in the same manner in the other crystal form (data not shown). This finding suggests that tetramer formation is not an artefact of crystal packing. The four protomers superimpose well, with an averaged rmsd value of 0.7 Å for comparable Cα atoms (Fig. 1B). This observation indicates that the structures of the four protomers are essentially identical, except for the N-terminal region of the β-chain.