Thus, the structure of the A molecule is hereafter considered to be representative of the Bg7S protomer, unless otherwise noted.

Bg7S adopts a β-rich structure with several α-helices (Fig. 1C). Bg7S is post-translationally cleaved between Ser251 and Ser252, resulting in the α-chain and β-chain. Although these chains are intricately folded, the structure of Bg7S is roughly divided into the α-domain and β-domain. Bg7S has 12 cysteines in positions similar to those found in the primary structures of other XEGIPs and TAXIs, and these residues form six disulfide bonds. Because Bg7S is secreted from seeds in response to various stresses, such as heat [15], these disulfide bonds supposedly stabilize the three-dimensional structure of Bg7S. The Cys209–Cys418 bond seems to be significant for stabilization in particular, because it links the α-chain and β-chain (Fig. 1C).

A search for homologous structures of Bg7S by DALI [16] revealed that the structure of Bg7S is similar to those of the xylanase inhibitor TAXI-IA [Protein Data Bank (PDB) ID 1T6G, *Z*-score = 39.7] (Fig. 1D) and aspartic proteases such as pepsin (PDB ID 1MPP, *Z*-score = 29.7). Structure-based sequence alignment indicated that secondary structural elements are well conserved between Bg7S and TAXI-IA, whereas deletions and insertions in some loop regions are observed (Fig. 2A). In addition, although TAXI-IA also has 12 cysteines forming disulfide

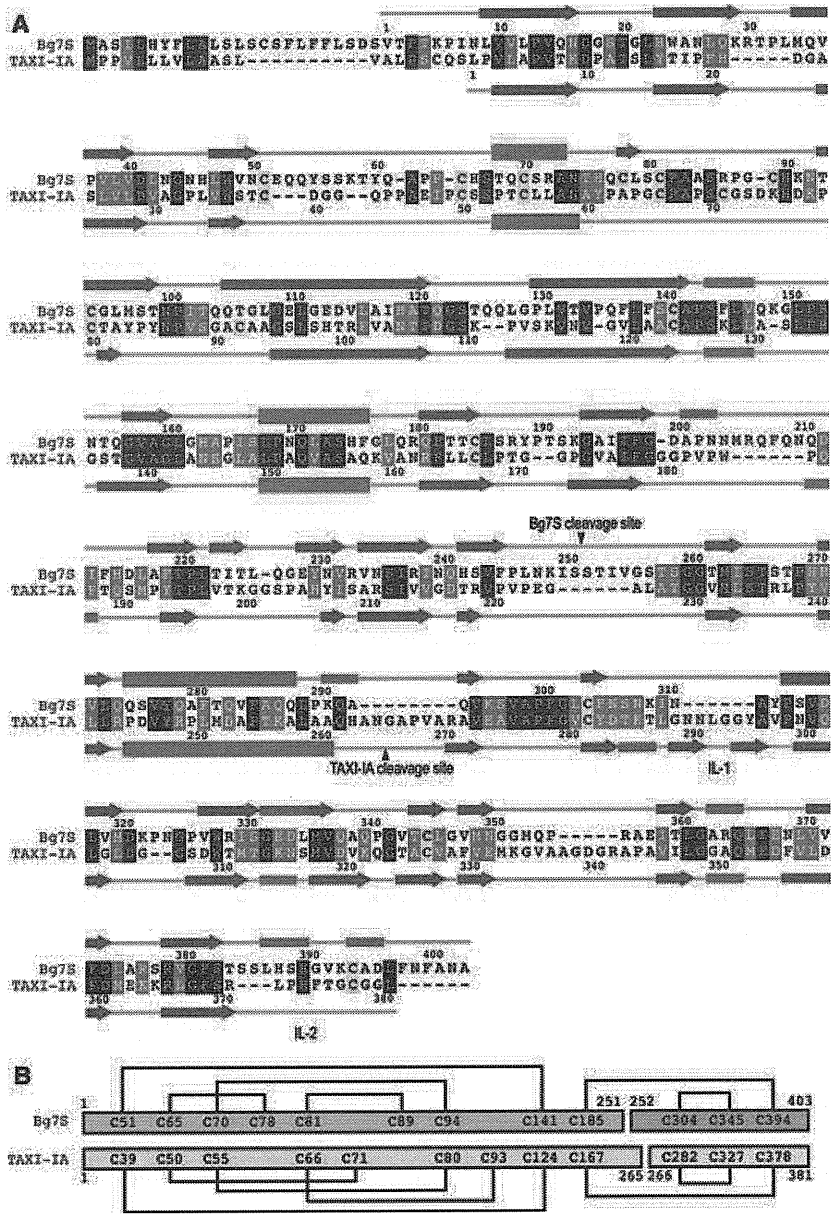



**Fig. 2.** Primary structures of Bg7S (soybean) and TAXI-IA (wheat). (A) Sequence alignment of Bg7S and TAXI-IA. Identical and homologous residues are highlighted by black and gray backgrounds, respectively. All cysteines are highlighted by a yellow background. Bg7S shares 26% amino acid identity with TAXI-IA. The secondary structures of Bg7S and TAXI-IA are shown above and below the sequences, respectively. The β-strand, α-helix and $3_{10}$-helix are shown in blue, red and magenta, respectively. (B) Disulfide bonds of Bg7S (upper) and TAXI-IA (lower).

bonds, the positions of the disulfide bonds in Bg7S are different from those in TAXI-IA (Fig. 2B) [7]. Both Bg7S and TAXI-IA adopt a pepsin fold. The pseudo-active site of Bg7S corresponding to pepsin is located in the cleft between the α-domain and β-domain, as observed in TAXI-IA [7] (Fig. 1C). However, both Bg7S and TAXI-IA lack protease activity, because one aspartate corresponding to the catalytic residue of pepsin is replaced by Ser265 and Ser235 in Bg7S and TAXI-IA, respectively.

## Assembly of Bg7S in solution

In marked contrast to TAXI-IA, Bg7S forms a tetramer with pseudo-*222* symmetry, as mentioned above. A number of water molecules are found in the protomer–protomer interfaces (Table 1), implying that assembly of Bg7S might be dynamically altered by solution conditions. To investigate the assembly of Bg7S in solution, we first performed an analytical ultracentrifugation (AUC) experiment, based on the sedimentation velocity method, at pH 7.4 (Fig. 3A). Sedimentation velocity analysis showed major and minor peaks corresponding to Bg7S tetramers and dimers, respectively. This observation indicates that there is an equilibrium between tetramers and dimers. Next, we performed sedimentation equilibrium analysis under the same buffer conditions (Fig. 3B). A tetramer–dimer self-association model was used for data analysis, and the dissociation constant ($K_d$) for dissociation of the Bg7S tetramer from the dimer was estimated to be 0.83 μM. We also performed size exclusion chromatography (SEC) to investigate the pH dependency of self-assembly of Bg7S (Fig. 3C). SEC analysis revealed the pH-dependent dynamic assembly of Bg7S in solution. At neutral pH (7.0), Bg7S formed a tetramer, a finding consistent with the results of AUC. In contrast, Bg7S was found to exist as a monomer at acidic pH (4.0). Interestingly, Bg7S seemed to form a dimer at both weakly acidic pH (6.0) and weakly basic pH (8–9).

**Table 1.** ΔASA in dimer formation. The ΔASA of the AB dimer is defined as [(ASA of A) + (ASA of B) − (ASA of AB dimer)]/2. The number of water molecules in the dimer interface was detected with ASV CALCULATOR [34]. ASA was calculated with a program kindly provided by M Maeda (National Institute of Agrobiological Sciences, Japan).

| | ΔASA (Å$^2$) | No. of water molecules |
|---|---|---|
| AB dimer | 1462 | 18 |
| BC dimer | 1493 | 24 |
| CD dimer | 1727 | 25 |
| DA dimer | 1511 | 22 |

Because structural analysis revealed that Bg7S forms a tetramer with pseudo-*222* symmetry (Fig. 1A), there are potentially two types of dimer formation, namely AB (or CD) and DA (or BC). The former and latter are designated face-to-face (FTF) and back-to-back (BTB) dimers, respectively. To assess which dimer is more plausible, the difference in accessible surface area (ΔASA) in each dimer was calculated (Table 1). It is conceivable that a dimer with larger ΔASA is more plausible. We found that the ΔASAs of the AB and DA dimers were comparable. Although the ΔASA of the CD dimer was slightly larger than the others, this was attributable to the N-termini of the β-chains (Fig. 1A). Those findings imply that both FTF and BTB dimers might be plausible. However, the electrostatic potential provided further insights into dimer formation (Fig. 3D, left panel). The FTF and BTB dimers utilize, respectively, acidic and basic surfaces during their formation. As a result, FTF and BTB dimers are supposed to be formed in weakly basic and weakly acidic conditions, respectively. Very recently, it has been reported that the formation of lupin γ-conglutin oligomers is dependent on pH [17]. γ-Conglutin undergoes a tetramer–dimer–monomer transition from neutral to acidic pH, which is consistent with our findings for Bg7S. Furthermore, Bg7S shares 63% amino acid identity with γ-conglutin. A homology model of γ-conglutin was build by SWISS-MODEL [18], using the structure of the Bg7S protomer as a template. In this homology model, the electrostatic potential of γ-conglutin is very similar to that of Bg7S (Fig. 3D, right panel). Thus, pH dependence of dynamic assembly might be a general feature of legume XEGIP proteins.

## Bg7S does not inhibit GH11 or GH12 enzymes

XEGIP was originally found to inhibit GH12 enzymes and not to inhibit GH11 enzymes. Thus, on the basis of this analogy with XEGIP, we first investigated whether or not Bg7S inhibits GH12 enzymes (Fig. 4A,B). Surprisingly, Bg7S did not inhibit either XEG or FI-CMC, a GH12 carboxymethyl cellulase from *A. aculeatus* [19]. We further investigated the activity of the GH11 xylanase ANXY in the presence of Bg7S (Fig. 4C). As expected, Bg7S did not affect the activity of ANXY. Even in the presence of leginsulin, a Bg7S-binding peptide, Bg7S did not inhibit GH12 or GH11 enzymes. Recently, it has been reported that lupin γ-conglutin does not inhibit GH12 endo-β-glucanase [20]. Therefore, we extracted XEGIPs from several legume seeds (azukibean, yardlongbean, and mungbean), and tested whether these proteins inhibited GH12 and GH11 enzymes (Fig. 4).

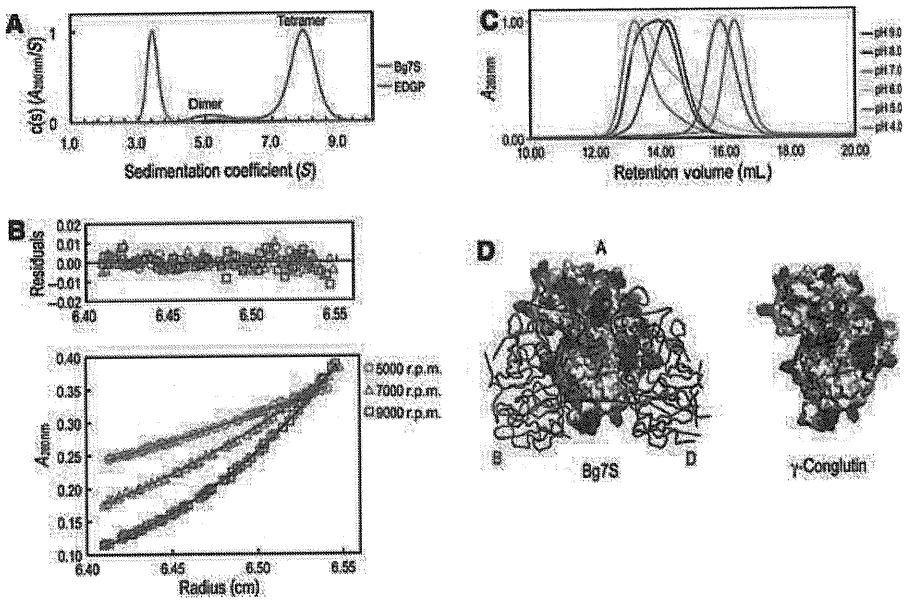



**Fig. 3.** Analysis of Bg7S assembly. (A) Sedimentation velocity analysis of Bg7S and EDGP. The sedimentation coefficient distributions of Bg7S and EDGP are indicated by the green and orange lines, respectively. EDGP is a monomeric standard. (B) Sedimentation equilibrium data are shown with the residuals from the best fit to a dimer–tetramer self-association model. Plots show data obtained at 5000 r.p.m. (red), 7000 r.p.m. (green), and 9000 r.p.m. (blue). (C) SEC elution profiles of Bg7S in various pH buffers are shown by the blue (9.0), light blue (8.0), green (7.0), yellow (6.0), red (5.0) and pink (4.0) lines. Absorbance at 280 nm is normalized. (D) Electrostatic potentials of the Bg7S A molecule (left) and the homology model of the γ-conglutin protomer (right). The blue and red surfaces indicate positive and negative potential, respectively. The B and D molecules of Bg7S are shown as loop representations. The colors of the B and D molecules of Bg7S correspond to those of Fig. 1A.

Like Bg7S, these legume XEGIPs did not affect the activities of GH12 and GH11 enzymes.

To date, structures of TAXI in complex with GH11 xylanase have been reported [7,9]. Structural superimposition of Bg7S on TAXI-IA in complex with ANXY (PDB ID 1T6G) provides significant insights into the structural basis of the lack of inhibition of GH11 enzymes by Bg7S (Fig. 5A). His374 and Leu292 of TAXI-IA, which are located in the loops termed, respectively, inhibition loop 2 (IL-2: residues 372–377) and inhibition loop 1 (IL-1: residues 290–294) in the present work, intrude into the active site of ANXY. His374 in IL-2 of TAXI-IA undergoes electrostatic interactions with the catalytic Glu79 and Glu170 of ANXY. In contrast, Leu292 in IL-1 of TAXI-IA undergoes a hydrophobic interaction with Tyr10 of ANXY. The interactions mimic those in the enzyme–substrate complexes (PDB ID 1BCX and 2QZ2) [7,9]. In addition, His374 of TAXI-IA interacts with Asp37 of ANXY. Bg7S lacks IL-1, and Leu292 of TAXI-IA is not conserved in Bg7S (Fig. 5A,B). Bg7S has His388 and His390 in IL-2 (residues 388–393). His390 is equivalent to His374 in IL-2 of TAXI-IA (Fig. 5B). However, the side chains of His388 and His390 do not face the protein exterior in the A molecule of the Bg7S tetramer. In the other protomers of the tetramer, the electron densities of IL-2 are ambiguous. This indicates that the IL-2 structure of Bg7S is potentially flexible, implying that these residues might interact with the catalytic residues of ANXY. However, sequence conservation in IL-2 between Bg7S and TAXI-IA is markedly lower than in any other region, and, furthermore, IL-2 of Bg7S is longer than that of TAXI-IA (Figs 2 and 5B). Thus, it is unlikely that IL-2 of Bg7S interacts with the active site.

The structure of XEGIP in complex with a GH12 enzyme has not been determined so far. However, the structures of both GH12 and GH11 enzymes adopt a similar β-jelly roll structure and have catalytic glutamates, indicating that Bg7S lacks inhibitory activity against GH12 enzymes for a similar reason. Recently, it has been reported that γ-conglutin, which also lacks IL-1, does not inhibit GH12 or GH11 enzymes [20]. Therefore, it is conceivable that legume XEGIPs in general do not inhibit either GH12 or GH11 enzymes.

## Conclusion

In this work, we have determined the crystal structure of Bg7S, which is the first three-dimensional structure

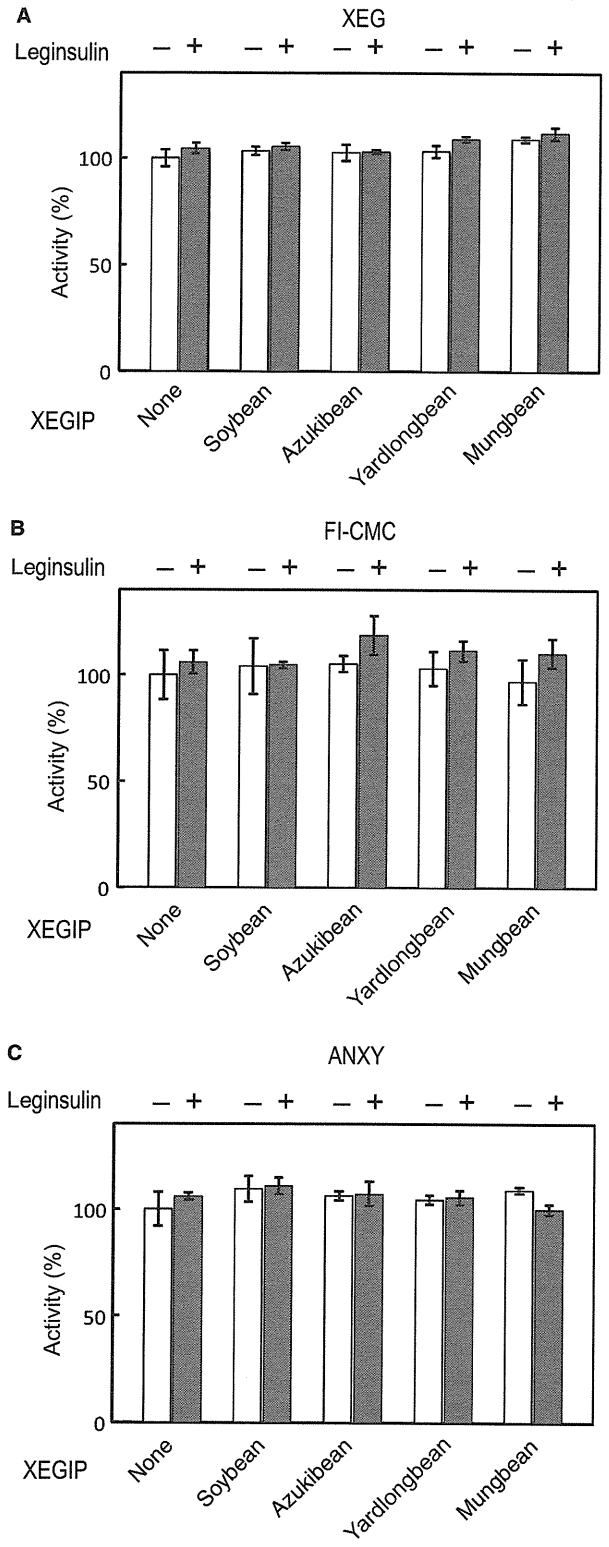



**Fig. 4.** Inhibitory activities of legume XEGIPs against GH12 and GH11 enzymes. The enzymatic activities of XEG (A), FI-CMC (B) and ANXY (C) in the presence of various legume XEGIPs were measured with or without the 4-kDa peptide from soybean (leginsulin).

of XEGIP. Bg7S forms a tetramer in a pH-dependent manner. Our biochemical characterization revealed that Bg7S, in contrast to XEGIP or TAXI, lacks inhibitory activity against both GH12 and GH11 enzymes. Furthermore, our study clarifies the structural basis for the lack of legume XEGIP inhibitory activity against both GH12 and GH11 enzymes. However, our results do not exclude the possibility that Bg7S functions as an inhibitory protein against GH enzymes other than GH12 and GH11 enzymes. The biochemical and biophysical features of legume XEGIPs are significantly distinct from those of XEGIPs from other plants. Thus, legume XEGIPs might be categorized differently from others. The physiological functions of legume XEGIPs, including Bg7S and γ-conglutin, remain unclear, and further functional studies are therefore required. Our structural and functional studies will provide significant clues for understanding the physiological function of legume XEGIPs, and will pave the way for future analysis.

## Experimental procedures

### Preparation and crystallographic analysis of Bg7S

Preparation and crystallization of the Bg7S have been described previously [14]. In brief, Bg7S was extracted from mature soybeen seeds (*Glycine max* L. Merrill cv. Miyagishirome). The protein was purified by using HisTrap Crude (GE Healthcare, UK Ltd, Little Chalfont, UK), HiTrap SP (GE Healthcare) and EconoPac CM (Bio-Rad Laboratory, Hercules, CA, USA) columns. The orthorhombic crystal was obtained by the hanging-drop vapor-diffusion method under the form II crystallization condition [14]. X-ray diffraction data were collected at Photon Factory beamline BL-5A, with a Quantum 315 CCD detector (Area Detector Systems, Corporation, San Diego, CA, USA) All diffraction data were processed with the HKL2000 [21]. The structure was solved by a molecular replacement method with MOLREP [22], using the crystal structure of EDGP (Yoshizawa *et al.*, unpublished work). Model building was performed with COOT [23]. Structure refinement was performed at 1.9-Å resolution with CNS [24] and REFMAC [25], and validated with PROCHECK [26]. The data collection and refinement statistics are given in Table 2.

### SEC and AUC experiments

SEC was performed with a Superdex 200 10/300 GL column (GE Healthcare). Bg7S was eluted with buffer solutions of various pH: 50 mM sodium acetate (pH 4.0, pH 5.0), 20 mM potassium phosphate (pH 6.0, pH 7.0), or 50 mM Tris/HCl (pH 8.0, pH 9.0), with 150 mM KCl.

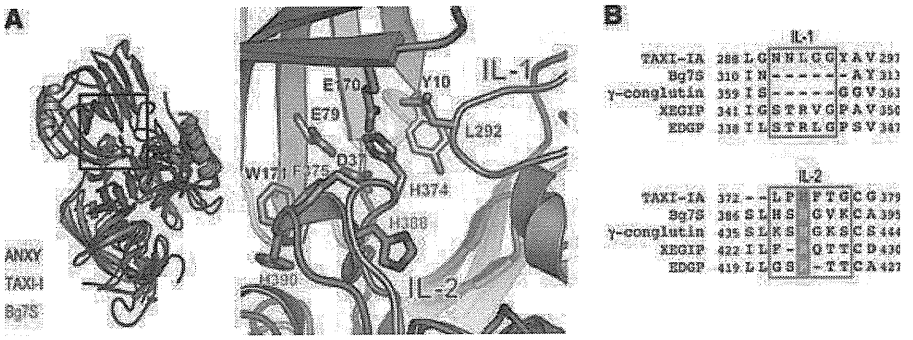



**Fig. 5.** Structural basis for the lack of inhibitory activity of Bg7S against GH12 and GH11 enzymes. (A) Structure of Bg7S superimposed on that of the TAXI-IA–ANXY complex (PDB ID 1T6G). The right panel shows a close-up view of the site of interaction between TAXI-IA and ANXY, roughly corresponding to the box in the left panel. Bg7S, TAXI-IA and ANXY are shown as green, light brown and gray ribbon representations, respectively. Residues that are significantly involved in the interaction between TAXI-IA and ANXY are shown as stick representations and labeled. His388 and His390 of Bg7S are also shown as stick representations. (B) Sequence alignment of IL-1 and IL-2 is shown in the upper and lower panels, respectively. IL-1 and IL-2 are indicated by light brown squares. Leu292 and His374 of TAXI-IA are highlighted in red. Homologous residues in IL-2 are highlighted by gray backgrounds.

AUC was performed with an Optima XL-I analytical ultracentrifuge (Beckman Coulter, Brea, CA, USA). The concentrations of the loaded protein solutions in the sedimentation velocity experiment were 0.88 mg·mL$^{-1}$ Bg7S or 0.91 mg·mL$^{-1}$ EDGP in a reference buffer (20 mM potassium phosphate, pH 7.4, and 250 mM KCl). EDGP was purified from carrot callus tissue [4]. Absorbance ($A_{280\ nm}$) scans were collected during sedimentation at 182 000 ***g***. Data analysis was performed with SEDIFIT [27,28] and SEDNTERP [29]. Sedimentation equilibrium experiments were performed in a six-channel centerpiece with quartz windows. The concentrations of the loaded protein solutions in the sedimentation equilibrium experiments were 0.18, 0.35 and 0.88 mg·mL$^{-1}$ in the reference buffer (20 mM potassium phosphate, pH 7.4, and 250 mM KCl). Data were obtained at 1820, 3562 and 5896 ***g***, respectively. Data analysis was performed by global analysis with ULTRASPIN (MRC Center for Protein Engineering, Cambridge, UK; http://www.mrc-lmb.cam.ac.uk/dbv/ultraspin2/).

## Preparation of XEGIPs from various legume seeds

Legume XEGIPs were purified from various dry mature seeds. We used soybean (*G.max* L. Merrill cv. Miyagishirome), yardolongbean (*Vigna unguiculata sesquipedalis* L. Verdc), azukibean (*Vigna angularis* L. cv. Dainagon), and mungbean (*Vigna radiata* R. Wilczek). For each, mature seeds were ground with water in a food processor (Cuisinart, Stamford, CT, USA) and a Polytron homogenizer (Kinematica, Bohemia, NY, USA), and then filtered through Miracloth (Merck KGaA, Darmstadt, Germany). The residue was stirred in buffer (20 mM potassium phosphate, pH 7.4, and 0.5 M NaCl) overnight at 4 °C, and then centrifuged at 43 667 ***g*** for 30 min. The supernatant contained mostly legume XEGIP, and was therefore used for enzyme inhibition assays. The purity of the proteins was checked by SDS/PAGE (Fig. S1).

**Table 2.** Data collection and refinement statistics. The values in parentheses are those for the highest-resolution shell (1.97–1.90 Å).

| | |
|---|---|
| Data collection | |
| Wavelength (Å) | 1.0000 |
| Space group | $P2_12_12$ |
| *a* (Å) | 135.2 |
| *b* (Å) | 161.2 |
| *c* (Å) | 84.8 |
| Resolution (Å) | 50.0–1.90 |
| Observed reflections | 720 554 |
| Unique reflections | 138 568 |
| *R*-merge (%) | 6.5 (36.5) |
| Completeness (%) | 95.6 (81.0) |
| $<I>/\sigma<I>$ | 11.3 (1.9) |
| Refinement | |
| Resolution (Å) | 20–1.91 |
| Refined reflections | 130 634 |
| Free reflections | 6532 |
| Protein atoms | 11 297 |
| Water molecules | 621 |
| *R* (%) | 21.1 |
| *R*-free (%) | 25.9 |
| rmsd | |
| Bond length (Å) | 0.018 |
| Bond angles (°) | 1.782 |
| Ramachandran plot | |
| Most favored (%) | 87.1 |
| Additional allowed (%) | 11.1 |
| Generously allowed (%) | 1.1 |
| Disallowed (%) | 0.7 |
| Averaged *B*-value (Å$^2$) | 47.07 |
| PDB code | 3AUP |

### Preparation of XEG, FI-CMC, and ANXY

cDNA encoding XEG, FI-CMC or ANXY was obtained by PCR-based gene synthesis. The oligonucleotides were designed by using DNAWORKS 3.1 [30] (http://helixweb.nih.gov/dnaworks/). The synthesized cDNAs were inserted into a pGEX6P-1 vector (GE Healthcare) at the *Bam*HI–*Xho*I site. The resulting plasmid encoded XEG, FI-CMC or ANXY with a glutathione-*S*-transferase (GST)-tag at the N-terminus. The expression vector was introduced into *Escherichia coli* BL21(DE3). The cells were grown at 37 °C to a cell density of 0.6–0.8 at 660 nm, and then for a further 6 h at 25 °C after the addition of 1 mM isopropyl thio-β-D-galactoside before being harvested. XEG and FI-CMC were purified by procedures similar to those already published [31,32]. In brief, XEG was purified with a glutathione Sepharose 4B (GS4B) resin (GE Healthcare), HiTrap Q HP column (GE Healthcare), and HiLoad Superdex 75 26/60 column (GE Healthcare). FI-CMC was purified with a GS4B resin (GE Healthcare) and HiLoad Superdex 75 26/60 column (GE Healthcare). The N-terminal GST tags of XEG and FI-CMC were cleaved by HRV3C protease, after affinity purification with GS4B (GE Healthcare). GST-fused ANXY was purified with GS4B resin (GE Healthcare). Because removal of the GST-tag of GST–ANXY reduced the stability of the protein, GST–ANXY was used in the following inhibition assay.

### Enzyme inhibition assay

The inhibitory activities of legume XEGIPs against GH enzymes were measured by the *p*-hydroxy-benzoic acid hydrazide method, where reducing sugar was detected by colorimetric reaction with *p*-hydroxy-benzoic acid hydrazide [33]. The assay for inhibition of XEG was performed in a 20-μL solution containing 50 mM sodium acetate (pH 4.6), 5 $mg{\cdot}mL^{-1}$ xyloglucan from tamarind seeds (DS Pharma, Osaka, Japan), 5 μg of legume XEGIP, and 100 ng of XEG. The assay for inhibition of FI-CMC was performed in a 50-μL solution containing 50 mM sodium acetate (pH 4.6), 5 $mg{\cdot}mL^{-1}$ carboxymethyl cellulose (Nacalai, Kyoto, Japan), 5 μg of legume XEGIP, and 100 ng of FI-CMC. The assay for inhibition of ANXY was performed in a 20-μL solution containing 50 mM sodium acetate (pH 4.6), 5 $mg{\cdot}mL^{-1}$ xylan (Sigma-Aldrich, St. Louis, MO, USA), 5 μg of legume XEGIP, and approximately 100 ng of GST–ANXY. In the assays in the presence of leginsulin, 0.5 μg of leginsulin was added to each reaction mixture including xyloglucan and XEGIP, and the solution was incubated for 10 min at room temperature. Then, each GH enzyme was added to the solution. The leginsulin used in the assay was chemically synthesized. The reaction mixtures were incubated at room temperature for 15 min, and the amount of reducing sugar was measured with a DU530 spectrometer (Beckman Coulter, Brea, CA, USA). The activity was measured at least three times for each sample. The average values are shown in Fig. 4.

### Figure preparation

Figures 1, 3D and 5A were prepared with PYMOL (http://www.pymol.org). All of the figures were modified with PHOTOSHOP and ILLUSTRATOR (Adobe Systems, San Jose, CA, USA).

## Acknowledgements

We acknowledge the kind support of the beamline staff of PF and SPring-8 for data collection. We also acknowledge the kind support of M Maeda (National Institute of Agrobiological Sciences, Japan) for calculation of ΔASA. This work was supported by KAKENHI (16770080, 17048023, and 19036025), the Protein 3000 Project and Target Protein Research Programs to M. Sato, T. Shimizu and H.H. from MEXT to M Sato, T Shimizu and H Hashimoto.

## References

1 Qin Q, Bergmann CW, Rose JK, Saladie M, Kolli VS, Albersheim P, Darvill AG & York WS (2003) Characterization of a tomato protein that inhibits a xyloglucan-specific endoglucanase. *Plant J* **34**, 327–338.
2 Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **280**(Pt 2), 309–316.
3 York WS, Qin Q & Rose JK (2004) Proteinaceous inhibitors of endo-beta-glucanases. *Biochim Biophys Acta* **1696**, 223–233.
4 Shang C, Sassa H & Hirano H (2005) The role of glycosylation in the function of a 48-kDa glycoprotein from carrot. *Biochem Biophys Res Commun* **328**, 144–149.
5 Naqvi SM, Harper A, Carter C, Ren G, Guirgis A, York WS & Thornburg RW (2005) Nectarin IV, a potent endoglucanase inhibitor secreted into the nectar of ornamental tobacco plants. Isolation, cloning, and characterization. *Plant Physiol* **139**, 1389–1400.
6 Debyser W, Derdelinkx G & Delcour JA (1997) Arabinoxylan solubilisation and inhibition of the barley malt xylanolytic system by wheat during mashing with wheat whole meal adjunct: evidence for a new class of enzyme inhibitors in wheat. *J Am Soc Brew Chem* **55**, 153–156.
7 Sansen S, De Ranter CJ, Gebruers K, Brijs K, Courtin CM, Delcour JA & Rabijns A (2004) Structural basis for inhibition of Aspergillus niger xylanase by Triticum aestivum xylanase inhibitor-I. *J Biol Chem* **279**, 36022–36028.
8 Fierens K, Gils A, Sansen S, Brijs K, Courtin CM, Declerck PJ, De Ranter CJ, Gebruers K, Rabijns A,

Robben J *et al.* (2005) His374 of wheat endoxylanase inhibitor TAXI-I stabilizes complex formation with glycoside hydrolase family 11 endoxylanases. *FEBS J* **272**, 5872–5882.
9 Pollet A, Sansen S, Raedschelders G, Gebruers K, Rabijns A, Delcour JA & Courtin CM (2009) Identification of structural determinants for inhibition strength and specificity of wheat xylanase inhibitors TAXI-IA and TAXI-IIA. *FEBS J* **276**, 3916–3927.
10 Yamauchi F, Sato K & Yamagishi T (1984) Isolation and partial characterization of a salt extractable globulin from soybean seeds. *Agric Biol Chem* **48**, 645–650.
11 Hanada K, Nishiuchi Y & Hirano H (2003) Amino acid residues on the surface of soybean 4-kDa peptide involved in the interaction with its binding protein. *Eur J Biochem/FEBS* **270**, 2583–2592.
12 Yamazaki T, Takaoka M, Katoh E, Hanada K, Sakita M, Sakata K, Nishiuchi Y & Hirano H (2003) A possible physiological function and the tertiary structure of a 4-kDa peptide in legumes. *Eur J Biochem/FEBS* **270**, 1269–1276.
13 Hanada K & Hirano H (2004) Interaction of a 43-kDa receptor-like protein with a 4-kDa hormone-like peptide in soybean. *Biochemistry* **43**, 12105–12112.
14 Yoshizawa T, Hashimoto H, Shimizu T, Yamabe M, Shichijo N, Hanada K, Hirano H & Sato M (2011) Purification, crystallization and X-ray diffraction study of basic 7S globulin from soybean. *Acta Crystallogr F Struct Biol Crystallogr Commun* **67**, 87–89.
15 Hirano H, Kagawa H & Okubo K (1992) Characterization of proteins released from legume seeds in hot water. *Phytochemistry* **31**, 731–735.
16 Holm L & Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545–W549.
17 Capraro J, Spotti P, Magni C, Scarafoni A & Duranti M (2010) Spectroscopic studies on the pH-dependent structural dynamics of gamma-conglutin, the blood glucose-lowering protein of lupin seeds. *Int J Biol Macromol* **47**, 502–507.
18 Arnold K, Bordoli L, Kopp J & Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics (Oxford)* **22**, 195–201.
19 Ooi T, Shinmyo A, Okada H, Hara S, Ikenaka T, Murao S & Arai M (1990) Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase) from Aspergillus aculeatus. *Curr Genet* **18**, 217–222.
20 Scarafoni A, Ronchi A & Duranti M (2010) gamma-Conglutin, the Lupinus albus XEGIP-like protein, whose expression is elicited by chitosan, lacks the typical inhibitory activity against GH12 endo-glucanases. *Phytochemistry* **71**, 142–148.
21 Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307–326.
22 Vagin A & Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30**, 1022–1025.
23 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
24 Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**(Pt 5), 905–921.
25 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255.
26 Laskowski RA, MacArthur MW, Moss DS & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.
27 Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* **78**, 1606–1619.
28 Schuck P, Perugini MA, Gonzales NR, Howlett GJ & Schubert D (2002) Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys J* **82**, 1096–1111.
29 Laue TM, Shah BD, Ridgeway TM & Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding SE, Rowe AJ & Horton LC, eds), pp. 90–125. Royal Society of Chemistry, Cambridge, UK.
30 Hoover DM & Lubkowski J (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res* **30**(10), e43.
31 Pauly M, Andersen LN, Kauppinen S, Kofod LV, York WS, Albersheim P & Darvill A (1999) A xyloglucan-specific endo-beta-1,4-glucanase from Aspergillus aculeatus: expression cloning in yeast, purification and characterization of the recombinant enzyme. *Glycobiology* **9**, 93–100.
32 Ohnishi A, Ooi T, Kinoshita S, Tomatsuri H, Umeda K, Ueda S, Hata Y & Arai M (1999) Analysis of a catalytic acidic pair in the active center of cellulase from Aspergillus aculeatus. *Biosci Biotechnol Biochem* **63**, 2157–2162.
33 Lever M (1972) A new reaction for colorimetric determination of carbohydrates. *Anal Biochem* **47**, 273–279.
34 Maeda MH & Kinoshita K (2009) Development of new indices to evaluate protein–protein interfaces: assembling space volume, assembling space distance, and global shape descriptor. *J Mol Graph Model* **27**, 706–711.

## Supporting information

The following supplementary material is available:
**Fig. S1.** SDS/PAGE of XEGIPs from various legume seeds.

This supplementary material can be found in the online version of this article.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

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

# Purification, crystallization and X-ray diffraction study of basic 7S globulin from soybean

**Takuya Yoshizawa,[a] Hiroshi Hashimoto,[a]* Toshiyuki Shimizu,[b] Mayuki Yamabe,[a] Naoki Shichijo,[a] Kazuki Hanada,[a] Hisashi Hirano[a] and Mamoru Sato[a]**

[a]Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro, Tsurumi, 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 4 October 2010
Accepted 9 November 2010



Basic 7S globulin (Bg7S) is expressed by soybeans in response to biotic or abiotic stress. Bg7S is capable of binding to a 4 kDa protein which is supposedly involved in cell proliferation. Bg7S is widely found not only in legumes, but also in other plants; however, its function is still unclear. Here, Bg7S was successfully crystallized. Orthorhombic and monoclinic crystals of Bg7S were obtained under different conditions and belonged to space groups $P2_12_12$, with unit-cell parameters $a$ = 111.9, $b$ = 130.1, $c$ = 287.8 Å, and $P2_1$, with unit-cell parameters $a$ = 85.3, $b$ = 137.6, $c$ = 162.1 Å, $\beta$ = 91.2°, respectively.

## 1. Introduction

Plants express various proteins in response to biotic or abiotic stress. When mature soybean seeds are immersed in hot water at 323–333 K, many proteins are released into the water. The major of these proteins is basic 7S globulin (Bg7S), which has a molecular mass of 43 kDa and consists of 27 and 16 kDa chains linked by disulfide bonds. Bg7S binds a 4 kDa protein from soybean called leginsulin, which is supposedly involved in cell proliferation. Thus, Bg7S is also called 43 kDa protein or leginsulin-binding protein (LBP; Hirano *et al.*, 1992). Normally, Bg7S is localized in plasma membranes and cell walls (Nishizawa *et al.*, 1994). Carrot callus tissue cultured with the 4 kDa protein grows more rapidly than in medium lacking the 4 kDa protein (Yamazaki *et al.*, 2003), suggesting that the 4 kDa protein may regulate growth, differentiation and cell proliferation. In previous work, the three-dimensional structure of the 4 kDa protein has been determined by NMR spectroscopy (Yamazaki *et al.*, 2003). The 4 kDa protein adopts a T-knot motif that relies on disulfide bonds to maintain its structure. The T-knot motif is shared by disulfide-rich small proteins with diverse functions, such as growth factors in animals, protease inhibitors, antimicrobial peptides in plants and toxin in insects (Lin & Nussinov, 1995). These small proteins bind to their target proteins to regulate or inhibit their functions.

Proteins that are homologous to the 4 kDa protein have only been discovered in legumes. In contrast, homologues of Bg7S are widely present in plants such as wheat, tomato and *Arabidopsis* (York *et al.*, 2004). It has been reported that several proteins that are homologous to Bg7S inhibit the activity of endoglucanases that belong to glycoside hydrolase families 11 or 12 (GH11 or GH12). GH11 and GH12 cleave xyloglucan or xylan, which are major components of the cell wall. It is therefore suggested that the homologues of Bg7S play important roles in plant defence by inhibition of the GH11 and GH12 endoglucanases secreted by invading phytopathogens. Functional studies of homologues from wheat (TAXI-I and TAXI-II; Gebruers *et al.*, 2004; Bourgois *et al.*, 2007), tomato (XEGIP; Qin *et al.*, 2003), tobacco (NEC4; Naqvi *et al.*, 2005), lupin ($\gamma$-conglutin; Scarafoni *et al.*, 2010) and carrot (EDGP; Shang *et al.*, 2005) have been reported. Of these, crystal structures of TAXI-I and TAXI-II, which specifically inhibit GH11 enzymes, have been determined (Sansen *et al.*, 2004; Pollet *et al.*, 2009). Bg7S is predicted to have diverse functions as a receptor-like protein in signal transduction in combination with the 4 kDa protein and as a defensive factor against plant-pathogen endoglucanases. However, the details of the function of Bg7S are still unclear. In the present paper, we describe the purification and crystallization of Bg7S and initial diffraction analysis of the crystals.

## 2. Materials and results

### 2.1. Preparation of Bg7S for crystallographic study

Soybean seeds (*Glycine max* L. Merrill cv. Miyagishirome) were ground in a food processor (Cuisinart) with 5 ml water per gram of soybean seeds at room temperature. The soybean paste was homogenized using a Polytron homogenizer (Kinematica) with an additional 5 ml water per gram of seeds at 277 K and then filtered using Miracloth (Merck Biosciences). The residue was further homogenized in 10 ml buffer (20 m*M* potassium phosphate pH 7.4 and 0.5 *M* NaCl) per gram of soybean seeds for 30 min at 277 K and centrifuged for 30 min at 277 K (43 667*g*). The supernatant was applied onto an Ni Sepharose column equilibrated with buffer consisting of 20 m*M* potassium phosphate pH 7.4 and 500 m*M* KCl. The bound proteins were eluted with a linear gradient from 0 to 100 m*M* imidazole. Fractions containing Bg7S were applied onto a HiTrap SP HP column (GE Healthcare) equilibrated with 20 m*M* potassium phosphate pH 7.4 and the bound proteins were eluted with a linear gradient from 0 to 400 m*M* KCl. Fractions containing Bg7S were applied onto an Econo-Pac CM column (Bio-Rad) equilibrated with 20 m*M* potassium phosphate pH 7.4 and the bound proteins were eluted with a linear gradient from 0 to 300 m*M* KCl. Fractions containing Bg7S were concentrated to 20 mg ml$^{-1}$ in 20 m*M* potassium phosphate pH 7.4 using Amicon Ultra 30 kDa cutoff filter units (Millipore). The purity of Bg7S was confirmed by SDS–PAGE with Coomassie Brilliant Blue stain. About 30 g mature soybean seeds yielded approximately 10 mg purified Bg7S.

### 2.2. Crystallization and initial crystallographic study of Bg7S

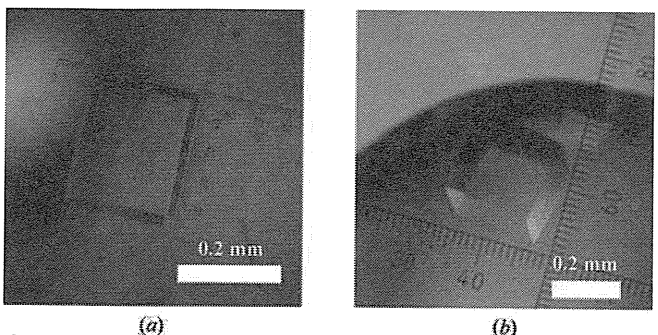
Screening of crystallization conditions for Bg7S was performed by the sitting-drop vapour-diffusion method using a Hydra II Plus One (Matrix) and 1150 different reservoir solutions from commercially available screening kits (Hampton Research, Molecular Dimensions, Emerald BioSystems and Qiagen). Crystals were obtained using several reservoir solutions, including Crystal Screen condition No. 42 (50 m*M* $KH_2PO_4$ and 20% PEG 800) and Wizard II condition No. 46 [1.0 *M* $(NH_4)_2HPO_4$, 0.1 *M* imidazole pH 8.0 and 0.2 *M* NaCl]. The buffer, the pH and the salt and precipitant concentrations of these crystallization conditions were optimized using the hanging-drop vapour-diffusion method. Two types of Bg7S crystal were obtained in different conditions (Figs. 1*a* and 1*b*). The orthorhombic crystal of Bg7S (form I) was obtained using a reservoir solution consisting of 50 m*M* $KH_2PO_4$, 100 m*M* sodium citrate pH 5.3, 6.5% PEG 8000 and 20% ethylene glycol after several days. Prior to the X-ray experiment, the form I crystal was transferred to a cryoprotectant solution consisting of 50 m*M* $KH_2PO_4$, 100 m*M* sodium citrate pH 5.3, 6.5% PEG 8000 and 30% ethylene glycol with a nylon loop and then cooled in a $N_2$-gas stream at 100 K. X-ray diffraction data were collected using a Quantum 315 CCD detector (Area Detector Systems Corp.) on beamline BL-41XU at SPring-8. Diffraction data were integrated, scaled and averaged with *HKL*-2000 (Otwinowski & Minor, 1997). The form I crystal belonged to space group $P2_12_12$, with unit-cell parameters $a$ = 111.9, $b$ = 130.1, $c$ = 287.8 Å. Data-collection statistics are summarized in Table 1. The asymmetric unit of the form I crystal is estimated to contain 8–12 molecules ($V_M$ = 3.04–2.02 Å$^3$ Da$^{-1}$). The monoclinic crystal (form II) was obtained using reservoir solution consisting of 1.0 *M* $(NH_4)_2HPO_4$ and 100 m*M* imidazole pH 8.0 after one week. For cryoprotection, the crystal was transferred into a cryoprotectant solution consisting of 1.0 *M* $(NH_4)_2HPO_4$, 100 m*M* imidazole pH 8.0 and 2.5 *M* $K_2HPO_4$. X-ray diffraction data were collected from the form II crystal using a Quantum 315 CCD detector (Area Detector Systems Corp.) on beamline BL-5A at the Photon Factory (PF). The form II crystal belonged to space group $P2_1$, with unit-cell parameters $a$ = 85.3, $b$ = 137.6, $c$ = 162.1 Å, $\beta$ = 91.2°. The asymmetric unit of the form II crystal is estimated to contain 8–10 molecules ($V_M$ = 2.96–2.15 Å$^3$ Da$^{-1}$). The preparation of heavy-atom derivatives and structure determination of Bg7S by the isomorphous replacement method are now in progress.

**Table 1**
Data-collection statistics for Bg7S crystals.

Values in parentheses are for the highest resolution shell.

| | Form I | Form II |
|---|---|---|
| Wavelength (Å) | 1.0000 | 1.0000 |
| Resolution range (Å) | 50.0–2.90 (3.00–2.90) | 50.0–2.60 (2.69–2.60) |
| Measured reflections | 643183 (38025) | 364026 (36799) |
| Unique reflections | 90761 (6671) | 105418 (10514) |
| Multiplicity | 7.1 (5.7) | 3.5 (3.4) |
| Completeness (%) | 96.2 (71.7) | 88.7 (89.0) |
| Mean $I/\sigma(I)$ | 13.9 (5.4) | 12.3 (3.3) |
| $R_{merge}$† (%) | 8.8 (30.2) | 10.1 (43.5) |

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




(*a*) (*b*)

**Figure 1**
Crystal forms I (*a*) and II (*b*) of Bg7S.

We acknowledge the kind support of the beamline staff of 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 Target Protein Research Program to MS, TS and HH from MEXT. We thank Professor J. R. H. Tame, Division of Protein Design Laboratory, Graduate School of Nanobioscience, Yokohama City University for English corrections.

## References

Bourgois, T. M., Nguyen, D. V., Sansen, S., Rombouts, S., Beliën, T., Fierens, K., Raedschelders, G., Rabijns, A., Courtin, C. M., Delcour, J. A., Van Campenhout, S. & Volckaert, G. (2007). *J. Biotechnol.* **130**, 95–105.
Gebruers, K., Brijs, K., Courtin, C. M., Fierens, K., Goesaert, H., Rabijns, A., Raedschelders, G., Robben, J., Sansen, S., Sorensen, J. F., Van Campenhout, S. & Delcour, J. A. (2004). *Biochim. Biophys. Acta*, **1696**, 213–221.
Hirano, H., Kagawa, H. & Okubo, K. (1992). *Phytochemistry*, **31**, 731–735.
Lin, S. L. & Nussinov, R. (1995). *Nature Struct. Biol.* **2**, 835–837.
Naqvi, S. M., Harper, A., Carter, C., Ren, G., Guirgis, A., York, W. S. & Thornburg, R. W. (2005). *Plant Physiol.* **139**, 1389–1400.
Nishizawa, N. K., Mori, S., Watanabe, Y. & Hirano, H. (1994). *Plant Cell Physiol.* **35**, 1079–1085.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pollet, A., Sansen, S., Raedschelders, G., Gebruers, K., Rabijns, A., Delcour, J. A. & Courtin, C. M. (2009). *FEBS J.* **276**, 3916–3927.
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.

Scarafoni, A., Ronchi, A. & Duranti, M. (2010). *Phytochemistry*, **71**, 142–148.
Shang, C., Sassa, H. & Hirano, H. (2005). *Biochem. Biophys. Res. Commun.* **328**, 144–149.
Yamazaki, T., Takaoka, M., Katoh, E., Hanada, K., Sakita, M., Sakata, K., Nishiuchi, Y. & Hirano, H. (2003). *Eur. J. Biochem.* **270**, 1269–1276.
York, W. S., Qin, Q. & Rose, J. K. (2004). *Biochim. Biophys. Acta*, **1696**, 223–233.

JJCO Japanese Journal of Clinical Oncology

*Jpn J Clin Oncol 2011;41(7)847–853*
*doi:10.1093/jjco/hyr060*
Advance Access Publication 15 May 2011

# Combined Functional Genome Survey of Therapeutic Targets for Clear Cell Carcinoma of the Kidney

**Hideaki Ito[1,2], Kazufumi Honda[1], Reiko Satow[1], Eri Arai[3], Miki Shitashige[1], Masaya Ono[1], Tomohiro Sakuma[4], Shigeru Sakano[2], Katsusuke Naito[2,5], Hideyasu Matsuyama[2] and Tesshi Yamada[1,*]**

[1]Divisions of Chemotherapy and Clinical Research, National Cancer Center Research Institute, Tokyo, [2]Department of Urology, Graduate School of Medicine, Yamaguchi University, Ube, [3]Divisions of Molecular Pathology, National Cancer Center Research Institute, Tokyo, [4]BioBusiness Group, Mitsui Knowledge Industry, Tokyo, and [5]Department of Urology, Mine City Hospital, Mine, Japan

*For reprints and all correspondence: Tesshi Yamada, Division of Chemotherapy and Clinical Research, National Cancer Centre Research Institute, Tokyo 104-0045, Japan. E-mail: tyamada@ncc.go.jp

Received January 6, 2011; accepted April 9, 2011

**Objective:** Emerging molecular targeting therapeutics have been incorporated into the management of advanced renal cell carcinoma; however, their efficacy remains limited. The aim of this study was to catalog potential therapeutic target molecules for renal cell carcinoma.
**Methods:** We first selected genes up-regulated in clear cell renal cell carcinoma relative to surrounding normal kidney tissues in 10 patients (Study Cohort) using high-density exon arrays that detect all potential transcripts predicted in the human genome. The selected genes were subjected to independent validation in another set of 10 patients (Validation Cohort) using real-time reverse transcriptase polymerase chain reaction and functional screening using small interfering RNA in six clear cell renal cell carcinoma cell lines.
**Results:** We identified 164 genes whose expression was significantly elevated in clear cell renal cell carcinoma ($P < 0.0001$ [Student's *t*-test] and at least a 3-fold change in transcription signal). We finally extracted 33 genes required for maintaining cell proliferation in at least two clear cell renal cell carcinoma cell lines. The 33 genes included 13 genes known to be associated with the development/progression of renal cell carcinoma, including *CAIX* and *FLT-1*, confirming the robustness of the current strategy.
**Conclusions:** Through a combination of genome-wide expression and functional assays, we identified a set of genes with high potential as targets for drug development. This method is rapid and comprehensive and could be applied to the discovery of diagnostic biomarkers and therapeutic targets for cancers other than clear cell renal cell carcinoma.

*Key words: exon array – functional screening – molecular targeting therapy – renal cell carcinoma – small interfering RNA*



## INTRODUCTION

Renal cell carcinoma (RCC) accounts for 3–4% of all human malignancies and is the 10th leading cause of cancer-related death in men (1). A quarter of RCC patients present with locally advanced or metastatic RCC (2). Surgical resection of the affected kidney is the most reliable treatment modality for localized RCC, while a third of such patients experience recurrence. RCC is not generally sensitive to chemotherapy or radiotherapy, and although immunotherapy using interferon-α or interleukin-2 has some efficacy, the overall response rate is limited to 10–15% and the median survival period of patients with metastatic RCC remains about 13 months (2).

Clear cell carcinoma is the major histological type, accounting for ~70–80% of RCC. The majority of clear cell RCC (CCRCC) carries the somatic mutation of the von Hippel-Lindau (*VHL*) gene (3). The functional loss of the *VHL* gene induces the overexpression of the transcription factor hypoxia inducible factor 1 and its downstream target genes, vascular endothelial growth factors (*VEGF*), glucose transporter, platelet-derived growth factor (*PDGF*) and transforming growth factor-α (*TGF*-α) (4). In

addition, the activation of the mammalian target of rapamycin (mTOR) signaling pathway has also been documented in CCRCC (5). Thus, anti-VEGF antibody (bevacizumab), small-molecule multikinase inhibitors that inhibit VEGF receptors (sorafenib, sunitinib, pazopanib and axitinib) and mTOR inhibitors (everolimus and temsirolimus) were recently developed, and these have been reported to have a favorable safety profile and to prolong time to progression for advanced RCC patients (6,7). However, such novel agents never eradicate cancer cells, and offer only a limited benefit to patients in terms of survival period (6–8). These facts indicate that there is an urgent need for developing new therapeutics.

In the present study, we adopted an unbiased genomic approach to identify genes involved in the growth of CCRCC cells. Here, we report 33 potential therapeutic targets for CCRCC.

## PATIENTS AND METHODS

### Tissue Samples

Sufficient amounts of surgically resected tissue were obtained from 20 patients who underwent radical nephrectomy at the National Cancer Center Hospital (Tokyo, Japan) for the treatment of histopathologically confirmed CCRCC in 2005–2006 (Study Cohort, $n = 10$) and in 1999 (Validation Cohort, $n = 10$), and all patient submitted written informed consent. The clinicopathological characteristics of these 20 patients are shown in Table 1. All patients underwent chest X-ray, computed tomography scan and bone scan examinations, and were staged in accordance with the American Joint Committee on Cancer /International Union Against Cancer TNM classification, fifth edition (1997). None of the patients received any preoperative treatments. Following nephrectomy, tumor and surrounding non-tumorous kidney tissues were immediately frozen in liquid nitrogen. The protocol of this study was reviewed and approved by the ethics committee of the National Cancer Center.

### RNA Extraction

Approximately 50 mg of frozen tissue was homogenized in 0.5 ml TRIzol (Invitrogen, Carlsbad, CA), and total RNA was extracted according to the protocol of the supplier. RNA

**Table 1.** Clinicopathological characteristics of patients with clear cell renal cell carcinoma (CCRCC) in this study

| Age | Sex | Histology | Tumor size (mm) | Grade | v | pV | pT | pN | M (organ) | Stage |
|---|---|---|---|---|---|---|---|---|---|---|
| Study set ($n = 10$) | | | | | | | | | | |
| 51 | Male | Clear cell | 25 | 2 | 0 | 0 | 1a | 0 | 0 | 1 |
| 66 | Female | Clear cell | 60 | 1 | 1 | 1 | 3b | 0 | 0 | 3 |
| 55 | Female | Clear cell | 45 | 1 | 0 | 0 | 1b | 0 | 0 | 1 |
| 56 | Male | Clear cell | 80 | 3 | 1 | 1 | 3b | 0 | 0 | 3 |
| 77 | Male | Clear cell | 25 | 1 | 0 | 0 | 1a | 0 | 0 | 1 |
| 63 | Female | Clear cell | 95 | 1 | 1 | 0 | 3a | 0 | 0 | 3 |
| 78 | Male | Clear cell | 65 | 1 | 1 | 0 | 1b | 0 | 0 | 1 |
| 53 | Male | Clear cell | 28 | 1 | 0 | 0 | 1a | 0 | 0 | 1 |
| 55 | Male | Clear cell | 25 | 1 | 0 | 0 | 1a | 0 | 0 | 1 |
| 64 | Male | Clear cell | 50 | 1 | 0 | 0 | 1b | 0 | 0 | 1 |
| Validation set ($n = 10$) | | | | | | | | | | |
| 39 | Male | Clear cell | 40 | 1 | 0 | 0 | 1a | 0 | 0 | 1 |
| 45 | Male | Clear cell | 23 | 1 | 0 | 0 | 1a | 0 | 0 | 1 |
| 70 | Male | Clear cell | 55 | 1 | 0 | 0 | 1b | 0 | 0 | 1 |
| 68 | Male | Clear cell | 70 | 3 | 1 | 1 | 3b | 0 | 1 (Lung) | 4 |
| 42 | Male | Clear cell | 50 | 2 | 0 | 0 | 1b | 0 | 0 | 1 |
| 56 | Male | Clear cell | 48 | 3 | 0 | 0 | 1b | 0 | 1 (Bone) | 4 |
| 55 | Female | Clear cell | 13 | 3 | 0 | 0 | 1a | 0 | 1 (Liver/stomach) | 4 |
| 62 | Male | Clear cell | 75 | 3 | 0 | 0 | 1b | 0 | 0 | 1 |
| 49 | Male | Clear cell | 45 | 2 | 0 | 0 | 1b | 0 | 0 | 1 |
| 62 | Male | Clear cell | 90 | 2 | 0 | 0 | 3a | 0 | 0 | 3 |

v, vascular involvement; pV, renal vein tumor thrombus.

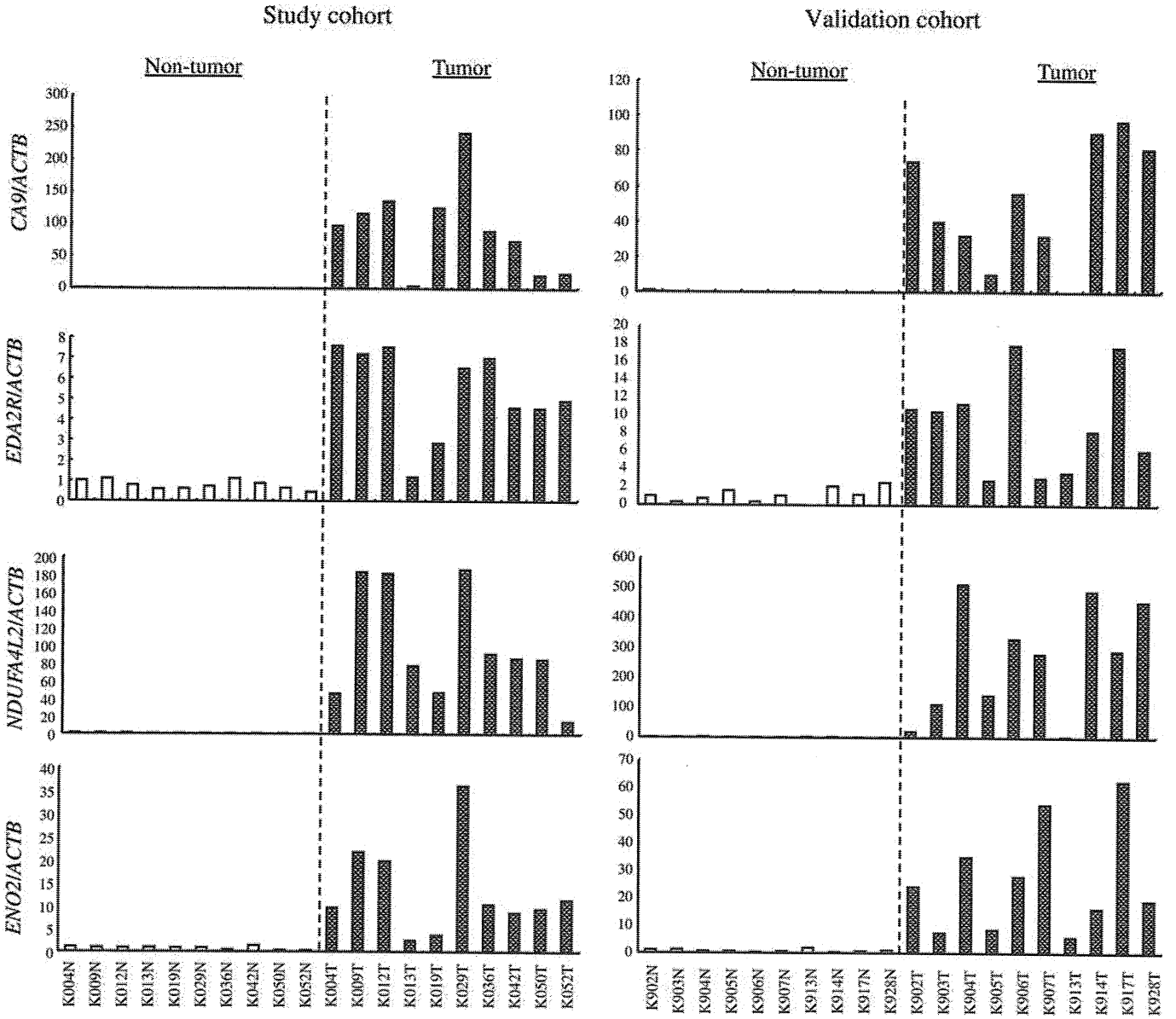



**Figure 1.** Real-time reverse transcription (RT) PCR analysis of *CAIX*, *EDA2R*, *NDUFA4L2* and *ENO2* genes in clear cell renal cell carcinoma (CCRCC) and surrounding renal cortex tissues.

quality was confirmed by electrophoresis, and only samples of intact RNA were used for further analysis.

## RCC Cell Lines

Six human cell lines (786-O, A-498, Caki-2, VMRC-RCW, Caki-1 and OS-RC-2) derived from RCC were used in this study. 786-O, A-498 and Caki-2 were obtained from the Cell Resource Center for Biomedical Research (Sendai, Japan). VMRC-RCW and Caki-1 were from the Japan Cancer Research Cell Bank (Osaka, Japan). OS-RC-2 was from RIKEN BRC cell bank (Tokyo, Japan). Caki-1, VMRC-RCW and A-498 were maintained in Eagle's minimum essential medium (Invitrogen) supplemented with 10% fetal bovine serum. 786-O and OS-RC-2 were maintained in RPMI 1640 medium (Invitrogen) supplemented with 10% fetal bovine serum. Caki-2 was maintained in McCoy's 5A medium (Invitrogen) supplemented with 10% fetal bovine serum.

## Exon Array Analysis

Total RNA (1 μg) was converted to end-labeled cRNA using a Whole Transcript Sense Target Labeling kit (Affymetrix, Santa Clara, CA) (9). Fluorescent cRNA probes were hybridized using Human Exon 1.0 ST arrays (Affymetrix), as instructed by the supplier (10). Data analysis for gene expression was carried out using the ArrayAssist software package (Version 5.5.1, Stratagene, La Jolla, CA). Antigenomic probes were used for background correction, and fluorescence intensity was normalized using the

**Table 2.** Biological process enrichment analysis

| | All genes | Selected genes | False discovery rate |
|---|---|---|---|
| Up-regulated | | | |
| GO:0051272_positive_regulation_of_cell_motion | 46 | 7 | 0.0178 |
| GO:0051270_regulation_of_cell_motion | 81 | 8 | 0.0451 |
| GO:0030335_positive_regulation_of_cell_migration | 42 | 6 | 0.0451 |
| Down-regulated | | | |
| GO:0007588_excretion | 41 | 13 | 1.08E–08 |
| GO:0006820_anion_transport | 49 | 11 | 1.74E–05 |
| GO:0006082_organic_acid_metabolic_process | 213 | 19 | 0.000359 |
| GO:0051179_localization | 1177 | 51 | 0.000553 |
| GO:0006811_ion_transport | 230 | 19 | 0.000553 |
| GO:0019752_carboxylic_acid_metabolic_process | 209 | 18 | 0.000553 |
| GO:0043436_oxoacid_metabolic_process | 209 | 18 | 0.000553 |
| GO:0046903_secretion | 168 | 16 | 0.000553 |
| GO:0006810_transport | 926 | 43 | 0.000554 |
| GO:0042180_cellular_ketone_metabolic_process | 214 | 18 | 0.000554 |
| GO:0051234_establishment_of_localization | 939 | 43 | 0.000702 |
| GO:0015698_inorganic_anion_transport | 25 | 6 | 0.00612 |
| GO:0015711_organic_anion_transport | 20 | 5 | 0.0248 |
| GO:0006519_cellular_amino_acid_and_derivative_metabolic_process | 121 | 11 | 0.0276 |
| GO:0006835_dicarboxylic_acid_transport | 5 | 3 | 0.0370 |

ExonRMA algorithm. Biological process enrichment analysis was performed using the GoMiner (11) with the Benjamini–Hochberg correction of the false discovery rate (FDR) for multiple tests. Only statistically significant processes showing an FDR $< 0.05$ were selected.

### Real-time Reverse Transcription PCR

First-strand cDNA was synthesized from 1 μg of total RNA using SuperScript reverse transcriptase (Invitrogen). Real-time PCR was performed using an ABI PRISM 7000 sequence detection system (Applied Biosystems, Foster City, CA) in accordance with the manufacturer's instructions. Relative quantification was achieved by normalization against the value for the β-actin (*ACTB*) gene, as described previously (12).

### Small Interfering RNA-based Functional Analysis

Three small interfering RNA (siRNA) duplexes for each of 168 genes were synthesized in a 96-well format (FlexiPlate siRNA, Qiagen, Valencia, CA) (13). The day before siRNA transfection, CCRCC cells were seeded at $2.5 \times 10^3$ cells per well in 96-well plates. Cells were then transfected with siRNA at a concentration of 10–50 nM using Lipofectamine 2000 (Invitrogen). Seventy-two hours after transfection, cell viability was quantified using CellTiter-Glo Luminescent Cell Viability Assay (Promega, Madison, WI). Luminescence was measured with a GloMax 96 Microplate Luminometer (Promega). A 10-fold serial dilution ($10^{-10}$ to $10^{-13}$ M) of standard adenosine triphosphate (ATP) (Promega) was included on each assay plate in order to generate a standard curve.

## RESULTS

### Gene Expression Analysis

Ten paired tissue samples of CCRCC and adjacent non-tumorous kidney (Study Cohort) were subjected to genome-wide expression analysis using GeneChip Human Exon 1.0 ST Array. The array contains ~5.4 million probes grouped into 1.4 million probe sets that can interrogate over 1 million exon clusters. The design of the probes is based upon not only mRNA sequences deposited in the RefSeq and GeneBank databases, but also upon all potential exon sequences deduced within the human genome. The exon array can detect mRNAs with low abundance as well as alternatively poly-adenylated and spliced mRNA, as the probes are designed to hybridize with the entire sequences of the predicted exons.

**Table 3.** List of selected genes by small interfering RNA (siRNA) functional screening

| Gene symbol | Gene description | Refseq ID | % Growth after siRNA transfection | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 786-O | A498 | Caki-2 | OS-RC-2 | VMRC-RCW | Caki-1 |
| *PGF* | Placental growth factor | NM_002632 | 82.81 | 60.6 | 87.38 | 98.08 | 112.26 | 75.82 |
| *PCTK3* | PCTAIRE protein kinase 3 | NM_002596 | 62.17 | 77.3 | 101.12 | 126.04 | 94.08 | 129.32 |
| *OLFML2A* | Olfactomedin-like 2A | NM_182487 | −12.42 | 69.8 | 83.27 | 90.59 | 115.21 | 54.15 |
| *HSF4* | Heat shock transcription factor 4 | NM_001040667 | 17.74 | 40.6 | 93.0 | 78.43 | 85.84 | 76.17 |
| *CP* | Ceruloplasmin (ferroxidase) | NM_000096 | 125.97 | 60.2 | 94.55 | 115.97 | 118.55 | 72.89 |
| *TRIM22* | Tripartite motif-containing 22 | NM_006074 | 90.51 | 57.5 | 72.36 | 100.48 | 91.62 | 90.61 |
| *SNORA16A* | Small nucleolar RNA, H/ACA box 16A | NR_002976 | 110.25 | 50.4 | 69.09 | 102.93 | 86.88 | 99.89 |
| *SLC2A1* | Solute carrier family 2 (facilitated glucose transporter), member 1 | NM_006516 | 78.02 | 76.4 | 95.27 | 91.84 | 102.61 | 98.04 |
| *SLC17A4* | Solute carrier family 17 (sodium phosphate), member 4 | NM_005495 | 77.21 | −1.8 | 29.34 | 47.50 | 36.97 | −14.40 |
| *SCARB1* | Scavenger receptor class B, member 1 | NM_005505 | 106.60 | 44.0 | 67.91 | 53.45 | 80.77 | 82.45 |
| *SAMD9L* | Sterile alpha motif domain containing 9-like | NM_152703 | 70.83 | 78.3 | 101.44 | 82.43 | 92.95 | 88.24 |
| *RGS5* | Regulator of G-protein signaling 5 | NM_003617 | 133.09 | 40.2 | 79.82 | 105.81 | 102.54 | 90.52 |
| *OSMR* | Oncostatin M receptor | NM_003999 | 80.18 | 78.7 | 108.50 | 94.53 | 117.78 | 105.64 |
| *NDUFA4L2* | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4-like 2 | NM_020142 | 60.57 | 68.6 | 99.13 | 90.10 | 103.96 | 89.64 |
| *MCAM* | Melanoma cell adhesion molecule | NM_006500 | 89.80 | 46.8 | 99.13 | 103.91 | 100.23 | 90.65 |
| *LOXL2* | Lysyl oxidase-like 2 | NM_002318 | 128.23 | 51.9 | 75.47 | 109.66 | 105.01 | 73.72 |
| *ITGA5* | Integrin, alpha 5 (fibronectin receptor, alpha polypeptide) | NM_002205 | 99.18 | 43.8 | 80.84 | 102.62 | 89.91 | 81.07 |
| *INHBB* | Inhibin, beta B | NM_002193 | 110.40 | 39.6 | 74.46 | 105.30 | 77.34 | 100.74 |
| *IGFBP3* | Insulin-like growth factor-binding protein 3 | NM_001013398 | 91.90 | 42.4 | 95.23 | 74.57 | 93.22 | 96.94 |
| *FLT1* | Fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor) | NM_002019 | 97.02 | 55.7 | 80.36 | 53.91 | 71.85 | 98.37 |
| *FCGR3A* | Fc fragment of IgG, low affinity IIIa, receptor (CD16a) | NM_000569 | 116.48 | 58.5 | 91.02 | 67.81 | 89.57 | 101.89 |
| *FABP7* | Fatty acid-binding protein 7, brain | NM_001446 | 127.85 | 54.1 | 74.07 | 106.70 | 107.49 | 90.36 |
| *EGLN3* | Egl nine homolog 3 (*Caenorhabditis elegans*) | NM_022073 | 111.31 | 40.6 | 76.87 | 72.80 | 84.39 | 108.02 |
| *DIRAS2* | DIRAS family, GTP-binding RAS-like 2 | NM_017594 | 42.74 | 44.4 | 96.29 | 71.55 | 80.02 | 76.87 |
| *CSPG4* | Chondroitin sulfate proteoglycan 4 | NM_001897 | 70.02 | 79.4 | 107.48 | 84.97 | 94.61 | 113.09 |
| *CDH13* | Cadherin 13, H-cadherin (heart) | NM_001257 | 97.73 | 78.8 | 97.71 | 84.94 | 84.69 | 88.69 |
| *CD200* | CD200 molecule | NM_001004196 | 87.69 | 50.1 | 102.32 | 83.56 | 87.32 | 94.72 |
| *CALCRL* | Calcitonin receptor-like | NM_005795 | 116.11 | 44.5 | 76.32 | 111.27 | 96.58 | 83.13 |
| *CA9* | Carbonic anhydrase IX | NM_001216 | 7.22 | 10.6 | 14.84 | 60.00 | 48.64 | −11.00 |
| *BARX2* | BARX homeobox 2 | NM_003658 | 92.70 | 61.0 | 100.83 | 96.22 | 103.86 | 117.49 |
| *AHNAK2* | AHNAK nucleoprotein 2 | NM_138420 | 109.69 | 41.3 | 78.06 | 94.93 | 103.92 | 83.69 |
| *ADM* | Adrenomedullin | NM_001124 | 163.48 | 48.3 | 79.82 | 100.07 | 103.96 | 96.28 |
| *ADFP* | Adipose differentiation-related protein | NM_001122 | 110.35 | 53.4 | 64.83 | 67.70 | 81.81 | 89.59 |

GTP, guanosine triphosphate; IgG, immunoglobulin G; NADH, nicotinamide adenine dinucleotide.

We confirmed that the transcripts of 460 annotated genes were differentially expressed between the background (non-tumorous) kidney and CCRCC tissues with statistical significance [$P < 0.0001$ (Student's $t$-test) and at least a 3-fold change in transcription signal]. Of the identified 460 genes, 168 transcripts were up-regulated and 292 were down-regulated in CCRCC (Supplementary data, Tables S1 and S2).

### Validation of Up-regulated Genes by Real-time RT–PCR

Of the 168 up-regulated genes identified by the exon array analysis, *CAIX*, *EDA2R*, *NDUFA4L2* and *ENO2* were randomly selected for independent validation using real-time RT–PCR. The expression of these four genes was consistently elevated in 10 additional samples (Validation Cohort),

as well as in the original 10 samples (Study Cohort) of CCRCC, when compared with non-tumorous kidney tissues in the same patients, thus confirming the reproducibility of the exon array experiment (Fig. 1).

### Gene Ontology Analysis

The 168 up-regulated and 292 down-regulated genes in CCRCC were subjected to gene ontology analysis (Table 2). We noted significant enrichment of genes involved in cell motion and migration among the 168 gene up-regulated in CCRCC (FDR $< 0.05$).

### High-throughput siRNA-based Functional Screening

In order to identify genes essential for CCRCC cell proliferation and survival, six cell lines were transfected with siRNA duplexes (mixtures of three constructs per gene) targeting the 168 genes up-regulated in CCRCC on a one-gene-per-well basis. Seventy-two hours after transfection, the relative abundance of metabolically active cells was determined by quantifying the relative production of ATP. The mean ATP level in triplicate wells was normalized against those of serially diluted standard ATP and non-targeting (negative control) siRNA included on the same assay plates. We designed three siRNA constructs targeting different parts of the transcripts, and performed the transfection experiment at the lowest possible effective concentration in order to minimize off-target effects (14). We found that siRNA targeting 33 genes reduced the ATP production by more than 20% in at least two of the six RCC cell lines (Table 3).

## DISCUSSION

Several new molecular targeting drugs have been approved recently for the treatment of advanced RCC, and many other compounds are under preclinical or early clinical investigation (6,7). Some of these have been shown to significantly improve quality of life, delay progression and extend survival, but rarely cure RCC patients (6,7). It is therefore necessary to identify the molecules essential for RCC cell growth and develop therapeutic approaches targeting these molecules. To date, several groups have reported the microarray-based molecular profiles of RCC (15–20); however, there have been no genome-wide functional surveys of therapeutic targets for CCRCC.

The advent of synthetic RNA interfering libraries has been reported to make systemic functional screening possible, and genes that sensitize lung cancer cells to chemotherapeutic drugs and the genes required for proliferation and survival of several cancer cell lines (13,21,22) have been successfully identified. We previously demonstrated that the combination of genome-wide expression and functional screening could be applied to the identification of therapeutic targets in hepatocellular carcinoma (9). In the present study, we screened an siRNA library targeting all 168 genes up-regulated in CCRCC, and identified 33 genes required for maintaining the viability of CCRCC cells.

Of the 33 genes, 13 [*PGF*, *SLC2A1*, *SCARB1*, *RGS5*, *IGFBP3*, *FLT-1* (*VEGFR-1*), *FABP7*, *EGLN3*, *CDH13*, *CAIX*, *ADM*, *CD200* and *ADFP*)] have previously been reported to be associated with development, progression or prognosis of RCC (23–35). This supports the suitability of our method for selecting molecular target genes. An additional 10 genes (*CP*, *OSMR*, *MCAM*, *LOXL2*, *ITGA5*, *INHBB*, *FCGR3A*, *CSPG4*, *CALCRL* and *BARX2*) have been reported to be associated with other types of cancer (14, 36–43). These genes may also be promising as molecular targets in RCC. The remaining 10 genes have not been reported to be associated with the biology of any types of cancer, and may serve as novel drug targets for CCRCC.

In the present study, we used a comprehensive genomic approach to catalog genes involved in the viability of CCRCC cells. Our method is a rapid and comprehensive approach that could be applicable to the discovery of molecular markers and therapeutic targets for CCRCC. Furthermore, our results may help us to clarify the molecular mechanisms underlying CCRCC development and progression. However, our results allow only preliminary conclusions and are limited to the listing of candidate genes. Further experimental and clinical studies of the 33 genes are necessary.

## Supplementary data

Supplementary data are available at http://www.jjco.oxfordjournals.org.

## Funding

This work was supported by the Third-Term Comprehensive Control Research for Cancer and Research on Biological Markers for New Drug Development conducted by the Ministry of Health and Labor of Japan, and the Program for Promotion of Fundamental Studies in Health Sciences conducted by the National Institute of Biomedical Innovation of Japan.

## Conflict of interest statement

None declared.

## References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin* 2009;59:225–49. Epub 2009 May 27.
2. Cohen HT, McGovern FJ. Renal-cell carcinoma. *N Engl J Med* 2005;353:2477–90.
3. Kaelin WG, Jr. The von Hippel-Lindau tumor suppressor protein and clear cell renal carcinoma. *Clin Cancer Res* 2007;13:680s–4.

4. Linehan WM. Molecular targeting of VHL gene pathway in clear cell kidney cancer. *J Urol* 2003;170:593–4.
5. Robb VA, Karbowniczek M, Klein-Szanto AJ, Henske EP. Activation of the mTOR signaling pathway in renal clear cell carcinoma. *J Urol* 2007;177:346–52.
6. Costa LJ, Drabkin HA. Renal cell carcinoma: new developments in molecular biology and potential for targeted therapies. *Oncologist* 2007;12:1404–15.
7. Garcia JA, Rini BI. Recent progress in the management of advanced renal cell carcinoma. *CA Cancer J Clin* 2007;57:112–25.
8. Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet* 2008;372:449–56.
9. Satow R, Shitashige M, Kanai Y, Takeshita F, Ojima H, Jigami T, et al. Combined functional genome survey of therapeutic targets for hepatocellular carcinoma. *Clin Cancer Res* 2010;16:2518–28.
10. Bemmo A, Benovoy D, Kwan T, Gaffney DJ, Jensen RV, Majewski J. Gene expression and isoform variation analysis using Affymetrix exon arrays. *BMC Genomics* 2008;9:529.
11. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* 2003;4:R28.
12. Shitashige M, Naishiro Y, Idogawa M, Honda K, Ono M, Hirohashi S, et al. Involvement of splicing factor-1 in beta-catenin/T-cell factor-4-mediated gene transactivation and pre-mRNA splicing. *Gastroenterology* 2007;132:1039–54.
13. Yamaguchi U, Honda K, Satow R, Kobayashi E, Nakayama R, Ichikawa H, et al. Functional genome screen for therapeutic targets of osteosarcoma. *Cancer Sci* 2009;100:2268–74.
14. Cobellis L, Cataldi P, Reis FM, De Palo G, Raspagliesi F, Pilotti S, et al. Gonadal malignant germ cell tumors express immunoreactive inhibin/activin subunits. *Eur J Endocrinol* 2001;145:779–84.
15. Takahashi M, Rhodes DR, Furge KA, Kanayama H, Kagawa S, Haab BB, et al. Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc Natl Acad Sci USA* 2001;98:9754–9.
16. Lane BR, Li J, Zhou M, Babineau D, Faber P, Novick AC, et al. Differential expression in clear cell renal cell carcinoma identified by gene expression profiling. *J Urol* 2009;181:849–60.
17. Yao M, Tabuchi H, Nagashima Y, Baba M, Nakaigawa N, Ishiguro H, et al. Gene expression analysis of renal carcinoma: adipose differentiation-related protein as a potential diagnostic and prognostic biomarker for clear-cell renal carcinoma. *J Pathol* 2005;205:377–87.
18. Hirota E, Yan L, Tsunoda T, Ashida S, Fujime M, Shuin T, et al. Genome-wide gene expression profiles of clear cell renal cell carcinoma: identification of molecular targets for treatment of renal cell carcinoma. *Int J Oncol* 2006;29:799–827.
19. Boer JM, Huber WK, Sultmann H, Wilmer F, von Heydebreck A, Haas S, et al. Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array. *Genome Res* 2001;11:1861–70.
20. Higgins JP, Shinghal R, Gill H, Reese JH, Terris M, Cohen RJ, et al. Gene expression patterns in renal cell carcinoma assessed by complementary DNA microarray. *Am J Pathol* 2003;162:925–32.
21. Whitehurst AW, Bodemann BO, Cardenas J, Ferguson D, Girard L, Peyton M, et al. Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature* 2007;446:815–9.
22. Schlabach MR, Luo J, Solimini NL, Hu G, Xu Q, Li MZ, et al. Cancer proliferation gene discovery through functional genomics. *Science* 2008;319:620–4.
23. Brown LF, Berse B, Jackman RW, Tognazzi K, Manseau EJ, Dvorak HF, et al. Increased expression of vascular permeability factor (vascular endothelial growth factor) and its receptors in kidney and bladder carcinomas. *Am J Pathol* 1993;143:1255–62.
24. Domoto T, Miyama Y, Suzuki H, Teratani T, Arai K, Sugiyama T, et al. Evaluation of S100A10, annexin II and B-FABP expression as markers for renal cell carcinoma. *Cancer Sci* 2007;98:77–82.
25. Fujita Y, Mimata H, Nasu N, Nomura T, Nomura Y, Nakagawa M. Involvement of adrenomedullin induced by hypoxia in angiogenesis in human renal cell carcinoma. *Int J Urol* 2002;9:285–95.
26. Furuya M, Nishiyama M, Kimura S, Suyama T, Naya Y, Ito H, et al. Expression of regulator of G protein signalling protein 5 (RGS5) in the tumour vasculature of human renal cell carcinoma. *J Pathol* 2004;203:551–8.
27. Hintz RL, Bock S, Thorsson AV, Bovens J, Powell DR, Jakse G, et al. Expression of the insulin like growth factor-binding protein 3 (IGFBP-3) gene is increased in human renal carcinomas. *J Urol* 1991;146:1160–3.
28. Krieg M, Haas R, Brauch H, Acker T, Flamme I, Plate KH. Up-regulation of hypoxia-inducible factors HIF-1alpha and HIF-2alpha under normoxic conditions in renal carcinoma cells by von Hippel-Lindau tumor suppressor gene loss of function. *Oncogene* 2000;19:5435–43.
29. Miki S, Matsumoto A, Nakamura Y, Itakura H, Kodama T, Yamamoto M, et al. Expression of scavenger receptors on renal cell carcinoma cells *in vitro*. *Biochem Biophys Res Commun* 1992;189:1323–8.
30. Morris MR, Hesson LB, Wagner KJ, Morgan NV, Astuti D, Lees RD, et al. Multigene methylation analysis of Wilms' tumour and adult renal cell carcinoma. *Oncogene* 2003;22:6794–801.
31. Tsuchiya N, Sato K, Akao T, Kakinuma H, Sasaki R, Shimoda N, et al. Quantitative analysis of gene expressions of vascular endothelial growth factor-related factors and their receptors in renal cell carcinoma. *Tohoku J Exp Med* 2001;195:101–13.
32. Yao M, Huang Y, Shioi K, Hattori K, Murakami T, Nakaigawa N, et al. Expression of adipose differentiation-related protein: a predictor of cancer-specific survival in clear cell renal carcinoma. *Clin Cancer Res* 2007;13:152–60.
33. Siva A, Xin H, Qin F, Oltean D, Bowdish KS, Kretz-Rommel A. Immune modulation by melanoma and ovarian tumor cells through expression of the immunosuppressive molecule CD200. *Cancer Immunol Immunother* 2008;57:987–96.
34. Amatschek S, Koenig U, Auer H, Steinlein P, Pacher M, Gruenfelder A, et al. Tissue-wide expression profiling using cDNA subtraction and microarrays to identify tumor-specific genes. *Cancer Res* 2004;64:844–56.
35. Bui MH, Seligson D, Han KR, Pantuck AJ, Dorey FJ, Huang Y, et al. Carbonic anhydrase IX is an independent predictor of survival in advanced renal clear cell carcinoma: implications for prognosis and therapy. *Clin Cancer Res* 2003;9:802–11.
36. Kunapuli SP, Singh H, Singh P, Kumar A. Ceruloplasmin gene expression in human cancer cells. *Life Sci* 1987;40:2225–8.
37. Lonnroth C, Andersson M, Arvidsson A, Nordgren S, Brevinge H, Lagerstedt K, et al. Preoperative treatment with a non-steroidal anti-inflammatory drug (NSAID) increases tumor tissue infiltration of seemingly activated immune cells in colorectal cancer. *Cancer Immun* 2008;8:5.
38. Peinado H, Del Carmen Iglesias-de la Cruz M, Olmeda D, Csiszar K, Fong KS, Vega S, et al. A molecular role for lysyl oxidase-like 2 enzyme in snail regulation and tumor progression. *EMBO J* 2005;24:3446–58.
39. Savarese TM, Campbell CL, McQuain C, Mitchell K, Guardiani R, Quesenberry PJ, et al. Coexpression of oncostatin M and its receptors and evidence for STAT3 activation in human ovarian carcinomas. *Cytokine* 2002;17:324–34.
40. Sellar GC, Li L, Watt KP, Nelkin BD, Rabiasz GJ, Stronach EA, et al. BARX2 induces cadherin 6 expression and is a functional suppressor of ovarian cancer progression. *Cancer Res* 2001;61:6977–81.
41. Shih IM. The role of CD146 (Mel-CAM) in biology and pathology. *J Pathol* 1999;189:4–11.
42. Yu YH, Kuo HK, Chang KW. The evolving transcriptome of head and neck squamous cell carcinoma: a systematic review. *PLoS One* 2008;3:e3215.
43. Zhang X, Green KE, Yallampalli C, Dong YL. Adrenomedullin enhances invasion by trophoblast cell lines. *Biol Reprod* 2005;73:619–26.

Published OnlineFirst August 9, 2011; DOI:10.1158/1055-9965.EPI-11-0400

Cancer Epidemiology, Biomarkers & Prevention

*Research Article*

# Identification of Adipophilin as a Potential Plasma Biomarker for Colorectal Cancer Using Label-Free Quantitative Mass Spectrometry and Protein Microarray

Junichi Matsubara[1,6], Kazufumi Honda[1], Masaya Ono[1], Shigeki Sekine[2], Yoshinori Tanaka[7], Michimoto Kobayashi[7], Giman Jung[7], Tomohiro Sakuma[3], Shoji Nakamori[8], Naohiro Sata[11], Hideo Nagai[11], Tatsuya Ioka[9], Takuji Okusaka[4], Tomoo Kosuge[4], Akihiko Tsuchida[5], Masashi Shimahara[10], Yohichi Yasunami[12], Tsutomu Chiba[6], and Tesshi Yamada[1]

## Abstract

**Background:** The aim of this study was to identify a new plasma biomarker for use in early detection of colorectal cancer.

**Methods:** Using the combination of hollow fiber membrane (HFM)-based low-molecular weight protein enrichment and two-dimensional image converted analysis of liquid chromatography and mass spectrometry (2DICAL), we compared the plasma proteome of 22 colorectal cancer patients with those of 21 healthy controls. An identified biomarker candidate was then validated in two larger cohorts [validation-1 ($n$ = 210) and validation-2 ($n$ = 113)] using a high-density reverse-phase protein microarray.

**Results:** From a total of 53,009 mass peaks, we identified 103 with an area under curve (AUC) value of 0.80 or higher that could distinguish cancer patients from healthy controls. A peak that increased in colorectal cancer patients, with an AUC of 0.81 and $P$ value of 0.0004 (Mann–Whitney $U$ test), was identified as a product of the *PLIN2* gene [also known as perilipin-2, adipose differentiation-related protein (ADRP), or adipophilin]. An increase in plasma adipophilin was consistently observed in colorectal cancer patients, including those with stage I or stage II disease ($P < 0.0001$, Welch's $t$ test). Immunohistochemical analysis revealed that adipophilin is expressed primarily in the basal sides of colorectal cancer cells forming polarized tubular structures, and that it is absent from adjacent normal intestinal mucosae.

**Conclusions:** Adipophilin is a plasma biomarker potentially useful for the detection of early-stage colorectal cancer.

**Impact:** The combination of HFM and 2DICAL enables the comprehensive analysis of plasma proteins and is ideal for use in all biomarker discovery studies. *Cancer Epidemiol Biomarkers Prev; 20(10); 2195–203. ©2011 AACR.*

## Introduction

Colorectal cancer is the second leading cause of cancer deaths in Western countries (1) and is the third leading cause of cancer deaths in Japan, where there were more than 43,000 estimated colorectal cancer deaths in 2008 (2). Treatment of colorectal cancer without metastasis is relatively uncomplicated, and a favorable prognosis can be expected for these patients (3, 4). However, the 5-year survival rate of patients with metastatic colorectal cancer is estimated to be less than 5% (5), underscoring the importance of early detection. The modality used most commonly for colorectal cancer mass screening is fecal occult blood (FOB) test. Three large randomized trials showed that inclusion of FOB in colorectal cancer screening significantly reduces the rates of colorectal cancer mortality (6–8). However, FOB has a relatively high false positive rate (9, 10), and as a result, a large number of healthy individuals receive radiological or endoscopic

**Authors' Affiliations:** Divisions of [1]Chemotherapy and Clinical Research, and [2]Molecular Pathology, National Cancer Center Research Institute; [3]BioBusiness Group, Mitsui Knowledge Industry; [4]Hepatobiliary and Pancreatic Oncology and Hepatobiliary and Pancreatic Surgery Divisions, National Cancer Center Hospital; [5]Third Department of Surgery, Tokyo Medical University, Tokyo; [6]Department of Gastroenterology and Hepatology, Kyoto University Graduate School of Medicine, Kyoto; [7]New Frontiers Research Laboratories, Toray Industries, Kamakura; [8]Department of Surgery, Osaka National Hospital, National Hospital Organization; [9]Department of Hepatobiliary and Pancreatic Oncology, Osaka Medical Center for Cancer and Cardiovascular Diseases; [10]Department of Oral Surgery, Osaka Medical College, Osaka; [11]Department of Surgery, Jichi Medical University, Shimotsuke; and [12]Department of Regenerative Medicine and Transplantation, Fukuoka University Faculty of Medicine, Fukuoka, Japan

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

**Corresponding Author:** Junichi Matsubara, Department of Gastroenterology and Hepatology, Kyoto University Graduate School of Medicine, 54 Shogoinkawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan. Phone: 81-75-751-4302; Fax: 81-75-751-4303; E-mail: jmatsuba@kuhp.kyoto-u.ac.jp

**doi:** 10.1158/1055-9965.EPI-11-0400

Published OnlineFirst August 9, 2011; DOI:10.1158/1055-9965.EPI-11-0400



reexamination after the FOB test, placing excessive physical and physiologic burdens on examiners and examinees, as well as imposing an undue financial burden upon society. The only approved screening alternative to FOB for the diagnosis of colorectal cancer is testing for the tumor marker carcinoembryonic antigen (CEA). Unfortunately, CEA is not useful as a marker for the early detection of colorectal cancer (11). Therefore, it is necessary to identify a new biomarker to supplement these current diagnostic modalities.

Alterations in the protein content of clinical samples reflect the dynamic biological changes of patients more directly than changes in mRNA levels (12). Plasma/serum proteins are thus valuable resources for the discovery of biomarkers with direct clinical application. We previously developed a quantitative proteomics platform called 2-dimensional image converted analysis of liquid chromatography and mass spectrometry (2DICAL; ref. 13). This technology is especially advantageous in clinical studies in which a large number of patient samples must be compared. We were able to identify a number of plasma/serum biomarkers with high potential for clinical application using 2DICAL (14–18). However, the direct analysis of plasma/serum proteins using 2DICAL remains technically challenging. Proteins secreted by cancer cells are considerably diluted in the blood circulation and present only in a low concentration (19, 20). The concentration of serum/plasma proteins ranges over more than 10 orders of magnitude and thus the efficient removal of abundant plasma/serum proteins is essential for the detection of low-abundance cancer-related biomarker proteins (21).

In this study, we applied a high-performance hollow-fiber membrane (HFM) technology to the enrichment of low-molecular weight (LMW) proteins (17, 22) and searched for new plasma biomarkers that might be applicable to the early diagnosis of colorectal cancer. The LMW plasma protein fraction is made up of various functional proteins, such as cytokines, chemokines, and peptides and is considered to be a rich unexplored archive of biological information (20). The HFM-based technique (HFMT) utilizes a fully automated system that can separate and concentrate low-abundance plasma proteins from relatively high-molecular weight abundant proteins such as albumin, immunoglobulin, transferrin, and apolipoproteins with high efficiency and reproducibility (22). Here, we report the identification of adipophilin, an adipose differentiation–related protein, as a novel tumor marker for colorectal cancer through a comprehensive analysis of the LWM plasma proteome of colorectal cancer patients using HFM and 2DICAL technologies.

## Patients and Methods

### Plasma samples

Plasma samples were collected prospectively from 366 individuals and then split randomly into 3 cohorts [training, validation-1 (V1), and validation-2 (V2); Table 1]. The cohorts were essentially hospital based and consisted of healthy volunteers and newcomers (primarily to gastrointestinal services) between August 2006 and October 2008 at the following 7 hospitals in Japan: National Cancer Center Hospital (NCCH; Tokyo), Osaka National Hospital (ONH; Osaka), Jichi Medical School Hospital (JMS; Shimotsuke), Osaka Medical Center for Cancer and Cardiovascular Diseases (Osaka), Tokyo Medical University Hospital (TMUH; Tokyo), Osaka Medical College Hospital (OMC; Osaka), and Fukuoka University Hospital (Fukuoka). This multi-institutional collaborative study group was organized by the "Third-Term Comprehensive Control Research for Cancer" conducted by the Ministry of Health, Labour and Welfare of Japan and joined the International Cancer Biomarker Consortium (23). Written informed consent was obtained from every subject.

All patients diagnosed as having cancer had histologic or cytologic proof of colorectal adenocarcinoma. Demographic and laboratory data for the cases are summarized in Table 1. The staging of cancer was defined according to TNM classification by the International Union against Cancer (UICC). The Training cohort comprised 43 cases, including untreated colorectal cancer patients from TMUH ($n = 8$), JMS ($n = 9$), and ONH ($n = 5$), and healthy controls from NCCH ($n = 2$), TMUH ($n = 9$), OMC ($n = 6$), and ONH ($n = 4$). The V1 and V2 cohorts comprised 210 and 113 cases, respectively, from the 7 hospitals as described above. The V1 cohort included 101 patients with colorectal cancer and 109 healthy controls. The V2 cohort comprised 26 patients with colorectal cancer and 87 healthy controls.

For all the samples used in this study, the same protocol was used for blood collection, storage, and freeze/thawing to ensure absence of any preanalytical bias caused by differences in sample handling. Blood was collected in a tube with EDTA at the time of diagnosis. Plasma was separated by centrifugation and frozen at −80°C until analysis. Macroscopically hemolyzed samples were excluded from the present analysis. The protocol of this study was reviewed and approved by the institutional ethics committee board of each participating institute.

### Depletion of high-molecular weight plasma proteins

The plasma samples of the training cohort were filtered through a 0.22-μm pore size filter. Five hundred microliter of the sample was diluted by adding 3.5 mL 25 mmol/L ammonium bicarbonate buffer (pH 8.0). The total of 4 mL of the plasma dilution was injected into a HFMT machine (22). After 1 hour of fully automated operation, the solution containing LMW proteins was recovered and lyophilized.

### Liquid chromatography mass spectrometry

The HFMT-treated samples were digested with sequencing grade modified trypsin (Promega) and analyzed in duplicate using a nano flow high-performance