BmGlcDH is a member of the short-chain dehydrogenase/reductase (SDR) superfamily and has been used clinically to examine blood glucose levels. In addition, BmGlcDH is believed to have great potential in the regeneration of NAD(P)H for the biocatalytic synthesis of fine chemicals [1–3]. Two other types of GlcDHs have been used for blood glucose detection. One is a pyrroloquinoline-quinone-containing GlcDH (EC 1.1.5.2; PQQ-GlcDH), which is a homodimer with a subunit weight of ~ 50 kDa [4]. Although PQQ-GlcDH is self-sufficient and does not require additional cofactors, it lacks thermostability and shows broad substrate specificity toward various aldohexoses [5]. Some attempts have been made to overcome these problems by using directed evolution and site-directed mutagenesis [6,7]. Another GlcDH is an FAD-containing enzyme (EC 1.1.99.10; FAD-GlcDH) that has been isolated from two organisms, namely *Aspergillus oryzae* [8] and *Burkholderia cepacia* [9]; furthermore, the gene encoding FAD-GlcDH from *B. cepacia* has been cloned, and the enzyme has been characterized in detail. The *B. cepacia* FAD-GlcDH does not require additional cofactors and is relatively more thermostable than the PQQ-GlcDHs. However, the disadvantages of its relatively low catalytic activity at room temperature and its broad substrate specificity remain to be resolved [9]. Site-directed mutagenesis studies on FAD-GlcDH have been performed to improve its substrate specificity [10]. In contrast to these two types of GlcDHs, BmGlcDH requires the addition of NAD^+ to the assay mixture. However, BmGlcDH is currently applied for clinical assay because of its relatively narrow substrate specificity toward D-glucose [11–13].

BmGlcDH has the typical tetrameric structure of SDR enzymes [14] and is composed of identical subunits, each with a subunit mass of 28 kDa. So far, several BmGlcDHs have been characterized [11–13], and their corresponding genes have been cloned from various *Bacillus* species. These enzymes have a single Rossmann-fold domain known as an $NAD(P)^+$-binding motif. The active form of BmGlcDH is a homotetramer, which dissociates into inactive monomer or dimer at pH values above 9, in the absence of a high concentration of NaCl [15–17]. Several BmGlcDH mutants exhibiting enhanced thermostability at an alkaline pH and/or under low-salt conditions have been reported previously [18–22]. *B. megaterium* IAM1030 has four BmGlcDH isozymes (I–IV) that have been enzymatically characterized [14]. Among these four isozymes, BmGlcDH-IV has high thermostability and exceptionally high catalytic activity toward D-glucose. However, weak activity of this enzyme toward several other monosaccharides such as D-xylose, D-mannose, D-galactose and D-glucosamine has also been detected. More accurate measurements of blood glucose levels are therefore anticipated once the substrate specificity of BmGlcDH-IV is altered. Many structures of SDR enzymes have been reported in complexes with substrates/inhibitors such as sugars, steroids and alcohols [23–26], whereas no structural information for D-glucose binding is yet available. Herein, we have determined the X-ray structures of the substrate-free, NADH-bound and D-glucose-bound forms of BmGlcDH-IV. These structures reveal that the C-terminal carboxyl group derived from a neighboring subunit directly interacts with D-glucose. Based on the structures, we have performed site-directed mutagenesis of BmGlcDH-IV and have obtained a useful mutant that showed improvement of D-glucose specificity while retaining similar thermostability to the wild-type enzyme.

Results and Discussion

Structures of wild-type BmGlcDH-IV

All BmGlcDH-IV crystals in the ligand-free form, NADH-complex and D-glucose-complex belong to the same *C*2 space group. The structure of BmGlcDH-IV in complex with D-glucose was determined by the molecular replacement method. As a search model, a monomer of GlcDH from *Bacillus megaterium* IWG3 (BmGlcDH-IWG3; PDB code 1GCO) was used [27]. The structure of BmGlcDH-IV in ligand-free form and in complex with NADH was determined by the molecular replacement method with a monomer of D-glucose-bound BmGlcDH-IV as a search model. All models were refined to a resolution of 2.0 Å. A Ramachandran plot showed that almost 90% of residues were in the most favored region, while there were no residues in the disallowed region. The data collection and model refinement statistics are summarized in Table 1.

The asymmetric unit contains two BmGlcDH-IV monomers (subunits A and B) that are related by a non-crystallographic twofold symmetry. The homotetramer having a 222-point-group symmetry was generated from the dimer in the asymmetric unit with crystallographic twofold symmetry. The resultant tetrameric structure is shown in Fig. 1. No major conformational differences were observed between the ligand-free, NADH-bound and D-glucose-bound structures. The BmGlcDH-IV monomer comprises 261 amino acid residues and shares high sequence homology (82.4%) with BmGlcDH-IWG3, whose quaternary

Table 1. Crystallographic data. Data for the highest resolution shell are provided in parentheses. $R_{merge} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$ where $\langle I_h \rangle$ is the mean intensity of a set of equivalent reflections. $R_{work} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$ for the 95% of the reflection data used in the refinement. F_{obs} and F_{calc} are observed and calculated structure factor amplitudes, respectively. R_{free} is the equivalent of R_{work} except that it was calculated for a randomly chosen 5% test set excluded from the refinement.

	Wild-type BmGlcDH-IV (ligand-free form)	Wild-type BmGlcDH-IV (NADH complex)	Wild-type BmGlcDH-IV (D-glucose complex)	BmGlcDH-IV G259A mutant (ligand-free form)	BmGlcDH-IV A258F mutant (NADH-D-glucose ternary complex)
PDB code	3AUS	3AUT	3AUU	3AY7	3AY6
Data collection					
Beamline	AR-NE3A	AR-NE3A	AR-NW12A	AR-NE3A	BL-5A
Wavelength (Å)	1.0000	1.0000	1.0000	1.0000	1.0000
Space group	*C*2	*C*2	*C*2	*C*2	$P2_1$
Unit-cell parameters (Å, deg)	65.1, 128.5, 64.7, 90, 110.7, 90	65.0, 128.1, 64.7, 90, 110.4, 90	65.0, 128.0, 64.5, 90, 110.0, 90	65.1, 128.3, 64.6, 90, 110.7, 90	57.1, 127.0, 88.7, 90, 94.9, 90
Resolution range (Å)	50–2.00 (2.08–2.00)	50–2.00 (2.08–2.00)	50–2.00 (2.08–2.00)	50–1.90 (1.93–1.90)	50–2.10 (2.14–2.10)
Unique reflections	33 073	32 754	33 134	38 880	73 008
Redundancy	3.8 (3.4)	3.7 (3.3)	3.8 (3.8)	3.8 (3.7)	3.8 (3.8)
Completeness (%)	98.5 (92.1)	98.4 (89.5)	99.4 (99.9)	99.8 (98.8)	100.0 (100.0)
I/σ (I)	22.4 (4.9)	19.8 (3.9)	21.3 (4.7)	35.3 (3.7)	16.9 (2.9)
R_{merge}	0.050 (0.278)	0.064 (0.325)	0.049 (0.281)	0.050 (0.441)	0.099 (0.551)
Model refinement					
R_{work}	0.205	0.190	0.205	0.200	0.194
R_{free}	0.259	0.241	0.257	0.246	0.235
rmsd bond lengths (Å)	0.012	0.009	0.012	0.010	0.011
rmsd bond angles (deg)	1.19	1.19	1.20	1.14	1.29
Total atoms	4138	4200	4153	4164	8776
Average *B*-factors (Å^2)					
Overall	39.1	35.4	40.1	33.5	29.3
NADH	–	50.8	–	–	31.8
D-Glucose	–	–	45.5	–	24.3
Ramachandran plot					
Favored (%)	89.1	89.3	89.1	89.3	89.3
Allowed (%)	10.9	10.7	10.9	10.7	10.7
Outliers (%)	0.0	0.0	0.0	0.0	0.0

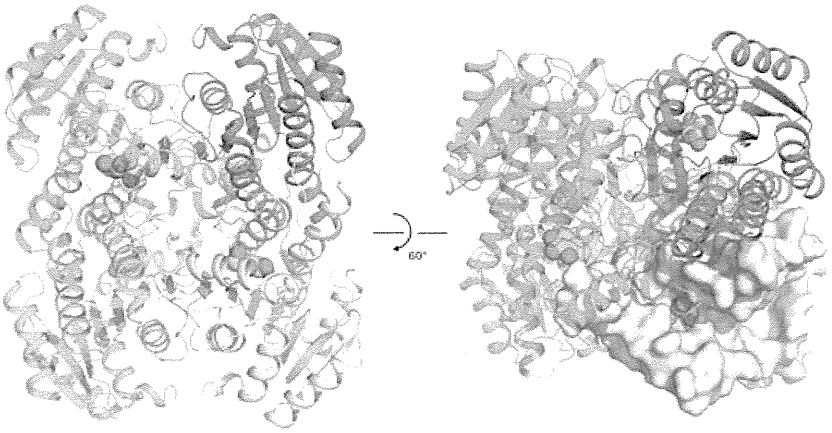

Fig. 1. Homotetrameric structure of BmGlcDH-IV. Each subunit comprising one tetramer is colored in yellow, green, cyan and magenta. The bound β-D-glucose is shown as a sphere. In the right panel, one of four subunits is represented as a surface model (light yellow) to clearly show the substrate-binding pocket and the C-terminal loop of the neighboring subunit (green) that directly interacts with β-D-glucose.

structure has been previously reported in D-glucose-free form [27]. The quaternary structure of BmGlcDH-IV in ligand-free form is very similar to that of BmGlcDH-IWG3, with an rmsd of 0.6 Å (for 261 Cα atoms). The BmGlcDH-IV subunit consists of seven-stranded central parallel β-sheets sandwiched by two arrays of three α-helices. BmGlcDH-IV represents a typical Rossmann fold in the core structure containing a GXXXGXG sequence, which is a characteristic fingerprint of the nucleotide-binding motif of the SDR enzyme [28,29].

Cofactor specificity

In our current structure of the NADH complex, an $F_o - F_c$ electron density map for NADH is clearly visible in chain A, but is rather ambiguous in chain B. Thus, the NADH model was built only in chain A. BmGlcDH-IV showed much lower K_M values for NAD^+ than $NADP^+$, while BmGlcDH-IWG3 showed much larger K_M values for NAD^+ than $NADP^+$, as has been previously reported by others [13,14]. Our current structure shows that the cofactor-binding pocket is very similar to that of BmGlcDH-IWG3 except for residues located near an adenine moiety. In BmGlcDH-IWG3, the side chain of the Arg39 residue in the β2α3 turn is probably important for the specificity of the $NADP^+$ molecule to form a salt bridge with the phosphate group of $NADP^+$. In contrast, there were no similar basic residues close to the 2′-phosphate group in BmGlcDH-IV. The corresponding residue is a Tyr39 that lies near the adenine ring, and probably plays a role in decreasing the enzyme's affinity toward $NADP^+$ (Fig. S1).

D-Glucose-binding mechanism

To understand the structural mechanism of D-glucose recognition of BmGlcDH-IV, we have determined the structure of BmGlcDH-IV in complex with D-glucose. Crystals of the D-glucose complex grew only when the crystallization solution contained both NADH and D-glucose. However, the electron density for NADH was completely missing. Steric hindrance between the reduced nicotinamide ring of NADH and the C1 hydrogen of D-glucose may stimulate the release of NADH as described below. The unambiguous electron density for the bound sugar indicated that the conformation of the pyranose ring is in a chair-type configuration (Figs 2 and S2). Amino acid residues forming hydrogen bonds with D-glucose are listed in Table S1, together with their bond distances. The C1 hydroxyl forms hydrogen bond interactions with the side chains of Tyr158 and Ser145. These residues are part of a catalytic triad (Tyr-Lys-Ser) in SDR enzymes. It is believed that Lys decreases the pK_a of the hydroxyl group in Tyr that acts as a catalytic base [30,31]. The Ser residue also functions in the stabilization of the enzyme–substrate complex during the reaction. In addition, clear electron density maps showed that the trapped D-glucose is in the β form with the C1 hydroxyl group in the equatorial configuration (Fig. S2). These results are consistent with previous reports that BmGlcDH acts on β-D-glucose, but not on α-D-glucose [12]. In addition, the side chains of Glu96 and Asn196 form hydrogen bonds with the C2 and C4 hydroxyl groups of β-D-glucose. The side chains of Lys199 and His147 also interacted with the C4 and C6 hydroxyl groups, respectively. It is interesting to note that the

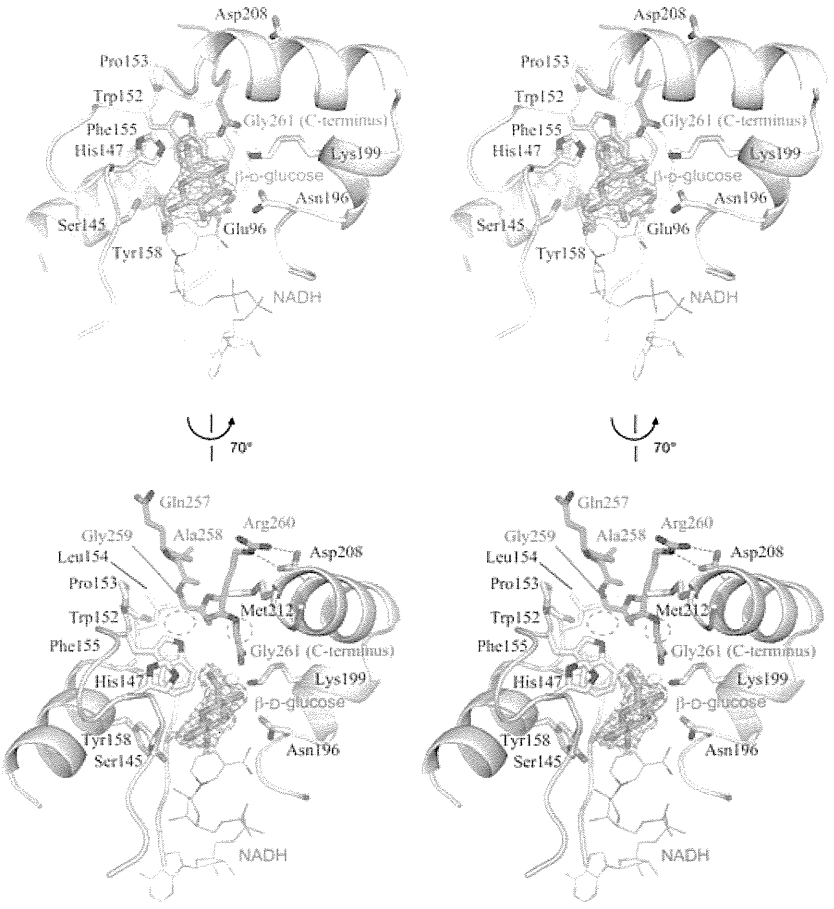

Fig. 2. Stereo-view showing the structure of the enzyme substrate-binding site in two different orientations. The bound β-D-glucose and the side chains of active-site residues are shown as stick models and labeled. A model of NADH is superimposed and shown as thin lines. Carbon, oxygen, nitrogen and phosphorus atoms of β-D-glucose and NADH are colored in cyan, red, blue and orange, respectively. The C-terminal loop of a neighboring subunit is shown in green. The red dotted circles in the lower panel represent the positions of the methyl group of the alanine mutants G259A and G261A. The methyl groups of Ala259 and Ala261 in the mutants are expected to cause steric hindrance with the main-chain oxygen of Pro153 and the side chain of Met212, respectively. The interactions between Arg260 and Asp208 are also shown as dotted lines in black. The $F_o - F_c$ omit electron density map for β-D-glucose is shown contoured at 3.0σ.

C-terminal carboxyl group (in Gly261) derived from a neighboring subunit is inserted into the active-site pocket and directly interacts via hydrogen bonds with C4 and C6 hydroxyl groups in the bound β-D-glucose. The side chain amino group of Lys199 is also located near the C-terminal carboxyl groups at a distance of < 3.0 Å. No crystal contacts are found in the vicinity of the substrate-binding pocket. These elaborate interatomic interactions probably play a critical role in stabilizing the conformation of the active-site and the bound β-D-glucose.

Since no structural differences were observed between the NADH and D-glucose complexes, it is possible to build a model of the putative ternary complex of BmGlcDH-IV-D-glucose-NADH by superimposing the NADH and D-glucose complexes. In the model, the C1 and C4 hydroxyl groups in the nicotinamide ring of NADH are in very close proximity (2.2 Å apart), suggesting that steric hindrance could have excluded the binding of NADH in the D-glucose-bound structure. The proximity between C4 in the nicotinamide ring and C1 in β-D-glucose also indicates that the binding position and orientation of β-D-glucose is not an artifact, and reflects the transient enzyme–substrate Michaelis intermediary complex that precedes hydride transfer from C1 in D-glucose to C4 in the nicotinamide ring.

Target residue selection for mutational analysis of BmGlcDH-IV C-terminal region

The C-terminal region of several SDR enzymes is reportedly involved in catalytic activity [32,33]. Our current structures of BmGlcDH-IV also reveal that the

C-terminal carboxyl directly interacts with substrate and stabilizes the structure of the active site (Fig. 2). To investigate the role of the C-terminal region in catalytic activity and substrate specificity, we attempted to disrupt the conformation of the C-terminal region in BmGlcDH-IV by using site-directed mutagenesis. Gly261 (at the C-terminus), Gly259 and Ala258 were chosen as target residues for mutational analysis, while Arg260 was ruled out because it is involved in the inter-subunit association through electrostatic interactions with Asp208 (Fig. 2). We have created a total of five single mutants as follows. For Gly261 and Gly259, we have created alanine and valine variants (G261A, G261V, G259A and G259V). Because there were no spaces for the methyl and isopropyl groups in the side chains of the variants (Fig. 2), these mutations are expected to disrupt the conformation of the C-terminal region and consequently affect the enzyme's activity and/or substrate specificity. In contrast, the methyl side chain in Ala258 is fully exposed to the solvent; thus, we created an A258F variant that is also expected to thermodynamically destabilize the C-terminal region. Additionally, we also created a C-terminal deletion mutant (ΔG261) to assess the importance of the C-terminal carboxyl group.

Comparison of enzymatic properties between BmGlcDH-III, BmGlcDH-IV and BmGlcDH-IV variants

First, we examined the pH dependence of the D-glucose oxidizing activity of BmGlcDH-III, -IV and the variants. Generally, wild-type BmGlcDH-III and BmGlcDH-IV show higher activity at alkaline pH because the proton-abstracting power of Tyr158 from the substrate is probably stronger at higher pH values [14]. However, the results showed that the optimal pH for five variants was around pH 6.0; furthermore, the activity drastically decreased at alkaline pH (Fig. 3). This might reflect a relaxation of the tetrameric assembly in the assay mixture at alkaline pH due to destabilization of the C-terminal region, which creates a part of the inter-subunit interface. However, gel filtration analysis suggested that both wild-type BmGlcDH-IV and its mutants exist as a tetramer in solution. Another possibility is that a different reaction pathway predominates for these mutants. Similar pH dependence was also reported with uronate dehydrogenase of the SDR family [34]; the authors proposed that the novel cationic intermediate does not require initial proton abstraction by the catalytic triad described above. The same reaction mechanism might also underlie BmGlcDH-IV oxidizing activity.

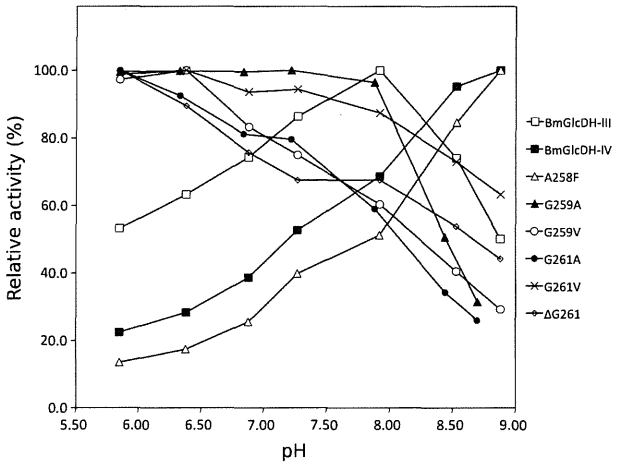

Fig. 3. The pH dependence at 37 °C of D-glucose oxidizing activity of BmGlcDH-III, -IV and its mutants.

Based on the results of pH dependence tests, the specific activities for D-glucose and other monosaccharides and the kinetic analysis for D-glucose were determined at both pH 6.0 and pH 8.0 (Tables 2 and 3). The specific activities of mutants G259V, G261A, G261V and ΔG261 for D-glucose were severely decreased at both pH 6.0 and pH 8.0. Kinetic analysis also indicated that K_M values for these mutants increased by 100–1000-fold, demonstrating the harmful influence of these mutations on enzyme performance (Table 2). However, the specific activity of the G259A mutant was slightly lower, but is still comparable with wild-type BmGlcDH-III and BmGlcDH-IV. Notably, the activity of G259A mutant with other analogous sugars was significantly decreased; in particular, its activity with D-xylose, D-mannose and D-galactose was not detected at pH 8.0. In contrast, the A258F mutation resulted in a slightly lowered activity but with considerably broadened substrate specificity. Kinetic analysis revealed that the K_M value for A258F with D-glucose slightly decreased to ~ 45% of the V_{max} value. We also examined the thermostability of wild-type BmGlcDH-III as well as BmGlcDH-IV and its variants by measuring its specific activity for D-glucose after incubation for 0–40 min at 65 °C (Fig. 4). The results clearly show that the thermostability of A258F and ΔG261 deteriorated markedly, and the level of activity was below that of BmGlcDH-III. The remaining activity of the G259V mutant was also lower than that of wild-type BmGlcDH-IV. The other three mutants retained almost the same level of remaining activity. The results suggest that the C-terminal carboxyl of Gly261 is important for both enzyme activity and thermostability. In summary, the G259A mutant

Table 2. Kinetic parameters of BmGlcDH-III, -IV and its mutants at 37 °C.

	pH 6.0		pH 8.0	
	K_M (mM)	V_{max} (U·mg^{-1})	K_M (mM)	V_{max} (U·mg^{-1})
BmGlcDH-III	11 ± 0.3	320 ± 11	63 ± 4.0	820 ± 34
BmGlcDH-IV	3.1 ± 0.3	230 ± 19	12 ± 0.8	810 ± 33
A258F	4.3 ± 0.2	110 ± 2.0	9.8 ± 1.0	310 ± 23
G259A	81 ± 7.1	480 ± 23	380 ± 16	960 ± 41
G259V	1.4×10^3 ± 96	510 ± 38	~ 4000[a]	~ 700[a]
G261A	4.1×10^2 ± 35	410 ± 25	~ 2000[a]	~ 800[a]
G261V	~ 4000[a]	~ 500[a]	~ 5000[a]	~ 600[a]
ΔG261	1.1×10^3 ± 170	260 ± 38	~ 2000[a]	~ 300[a]

[a] K_M and V_{max} were estimated by extrapolation with substrate concentrations ranging from 280 to 1400 mM.

Table 3. Comparison of substrate specificity of BmGlcDH-III, -IV and its mutants at 37 °C. ND, not detected. Values in parentheses refer to the specific activity (initial rate of enzyme reaction at each substrate concentration of 148 mM) (U·mg^{-1}).

	Substrate specificity (%)				
	D-glucose[a]	D-xylose	D-mannose	D-galactose	D-glucosamine
pH 6.0					
BmGlcDH-III	100 (308)	10.0	4.8	2.9	3.0
BmGlcDH-IV	100 (224)	35	18	9.1	11
A258F	100 (105)	92	35	9.5	30
G259A	100 (341)	1.7	ND	ND	ND
G259V	100 (51.1)	ND	ND	ND	ND
G261A	100 (108)	1.1	ND	ND	ND
G261V	100 (19.2)	1.5	ND	ND	ND
ΔG261	100 (32.2)	1.7	ND	ND	ND
pH 8.0					
BmGlcDH-III	100 (554)	4.0	1.7	1.1	7.5
BmGlcDH-IV	100 (811)	11	6.1	3.4	30
A258F	100 (319)	39	15	4.6	78
G259A	100 (256)	ND	ND	ND	2.5
G259V	100 (29.2)	ND	ND	ND	2.6
G261A	100 (62.7)	ND	ND	ND	1.9
G261V	100 (16.3)	ND	ND	ND	3.0
ΔG261	100 (21.8)	1.8	ND	ND	1.2

is the most useful mutant, exhibiting the narrowest substrate specificity whilst retaining its thermostability and specific activity at a practical level.

Structures of the G259A and A258F mutants

To understand the structural basis for the improved or broadened substrate specificity in the mutants, crystal structures of G259A and A258F mutant were also determined. The G259A crystals were obtained in the same space group as the wild-type BmGlcDH-IV under similar crystallization conditions with 4 mM NADH and 400 mM D-glucose. However, the electron density for D-glucose and NADH was missing. The electron density for the C-terminal residues 259–261 (chain A) and 260–261 (chain B) was also completely absent. These results suggest that these residues at the C-terminal region are protruded into the solvent region, and that the conformation of the C-terminal region is disordered (Fig. 5). Overall conformations except for these C-terminal residues are similar to that of wild-type BmGlcDH-IV, with rmsd of 0.2 Å for Cα atoms of the residues 1–258. We also determined the structure of the A258F mutant, which showed significantly broadened substrate specificity. A new crystal form belonging to the $P2_1$ space group was obtained under the same crystallization conditions as wild-type BmGlcDH-IV. The asymmetric unit contains one tetramer, and the four active sites were all occupied by both β-D-glucose and NADH. This is quite unexpected

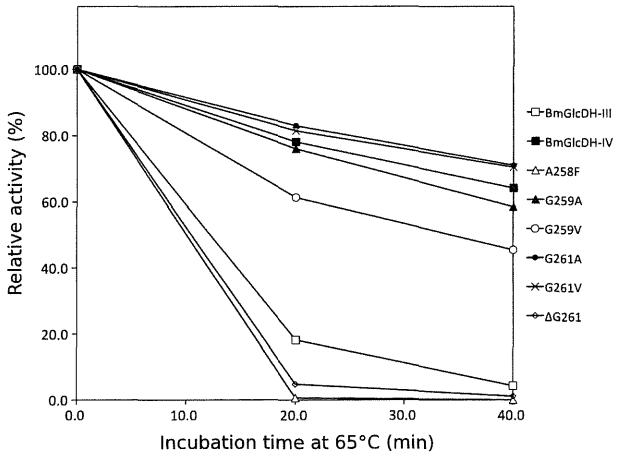

Fig. 4. Remaining activity of BmGlcDH-III, -IV and its variants after incubation at 0, 20 and 40 min at 65 °C. Assays were performed at 37 °C.

because NADH in the wild-type BmGlcDH-IV structure in complex with D-glucose is missing. Superimposing wild-type BmGlcDH-IV and the A258F mutant revealed no major conformational changes, although there was very slight movement of β-D-glucose toward Trp152 (< 0.4 Å) in the mutant structure to keep an appropriate distance between D-glucose and NADH. The distance between C1 in D-glucose and C4 in the nicotinamide ring is in the range of 3.1–3.3 Å. Two of four Phe258 residues in the tetramer are exposed to the solvent, while the other two Phe258 residues make a crystal contact with the Pro203-Val204 region of the symmetry-related neighboring tetramer. However, the crystal packing force does not influence the conformation of the C-terminal region, because the C-terminal residues including Phe258 in all subunits were well ordered in a similar manner to wild-type BmGlcDH-IV (Fig. 6).

The structural superimposition of all crystallographically independent BmGlcDH-IV subunits showed that one α-helix (αFG1; the nomenclature of the secondary structure elements follows the convention [14,25]) in the vicinity of the active site is rather mobile (Fig. 6). In particular, the residues of αFG1 in the structure of the G259A mutant have relatively high *B*-factor values (~ 60 Å^2) and move slightly away from the active site, probably due to the lack of electrostatic interactions between Arg257 and Asp208. The side chain amino group of Lys199 derived from the αFG1 makes hydrogen bonds to both O4 of D-glucose and the C-terminal carboxyl in the structure of wild-type enzyme and A258F mutant (Fig. 2). In the structure of G259A mutant, the side chain of Lys199 also moves away from the active site. The observation implies that the mobility of the C-terminal region and the αFG1 might be related to sugar-binding selectivity of the enzyme. Although the structures of the two mutants provided no clear explanation for the observed changes in the properties of the mutant enzyme, our mutational studies suggested that catalytic activity and substrate specificity could be accurately manipulated by introducing mutations at the C-terminal region. A more useful mutant might be generated using saturated mutagenesis on these C-terminal residues.

Conclusion

We have determined the structures of BmGlcDH-IV in ligand-free form, in complex with NADH and in complex with D-glucose. The structure of the D-glucose complex revealed that the C-terminal region of the polypeptide chain is inserted into the active-site pocket and that the C-terminal carboxyl group directly interacts with bound D-glucose. Mutational analyses demonstrated that disrupting the conformation of the C-terminal region greatly affected the catalytic activity,

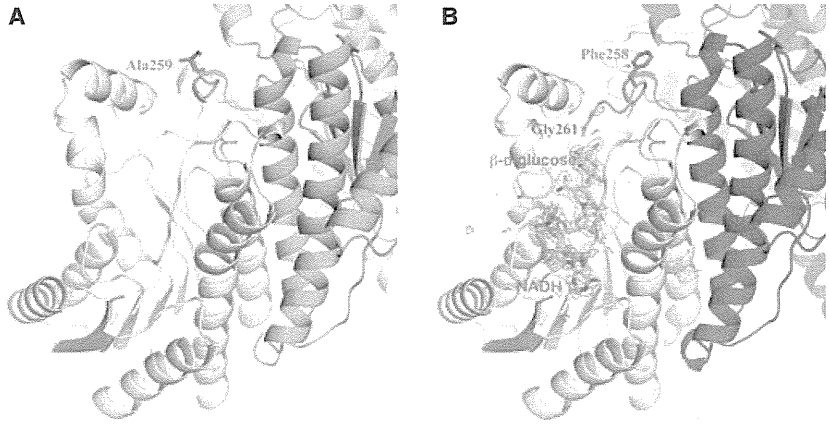

Fig. 5. The structure of NADH and the D-glucose-binding sites in mutants G259A (A) and A258F (B) using the same colors as applied to Fig. 1. For the A258F mutant, the $F_o - F_c$ electron density map for β-D-glucose and NADH is shown contoured at 3.0σ.

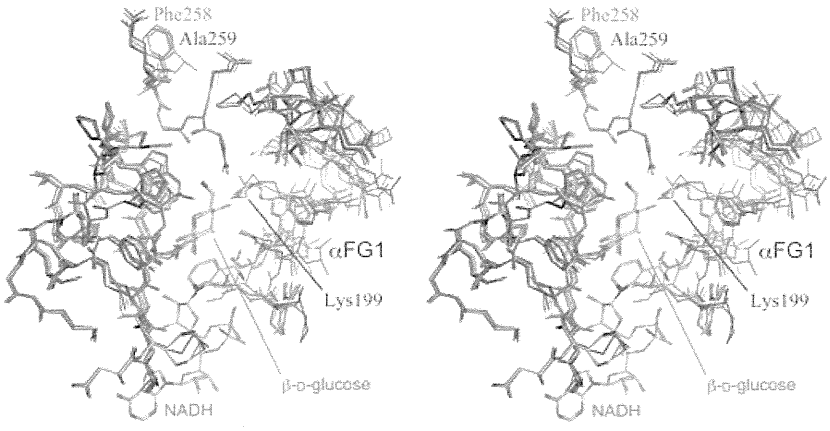

Fig. 6. Stereo-view superimposition of eight crystallographically independent active-site structures of BmGlcDH-IV. The wild-type BmGlcDH-IV complexed with β-D-glucose (two models, yellow), A258F mutant complexed with β-D-glucose and NADH (four models, red) and ligand-free G259A mutant (two models, blue) are shown. Carbon, oxygen, nitrogen and phosphorus atoms of β-D-glucose and NADH are colored in cyan, red, blue and orange, respectively. Mutated residues Phe258 and Ala259 are labeled. One of the substrate-binding residues Lys199 derived from a mobile α-helix (αFG1) is also labeled. This figure was prepared in the same orientation as in the lower panel of Fig. 2.

pH dependence and substrate specificity of the enzyme. Of the six mutants we created, the G259A mutant showed the most improved substrate specificity, while retaining its thermostability as well as a reasonable activity level against D-glucose. In some SDR enzymes such as GlcDHs and alcohol dehydrogenases, the C-terminal region is found to lie in close proximity to the active-site pocket and is important for the catalytic activity of the enzymes. Site-directed and/or saturated mutagenesis studies on the C-terminal loop in SDR enzymes might thus be a good strategy to obtain useful mutants having preferable substrate specificities.

Experimental procedures

Materials

Synthetic oligonucleotides were obtained from Hokkaido System Science Co. Ltd (Sapporo, Japan). All enzymes used for genetic manipulation were purchased from New England Biolabs Japan (Tokyo, Japan), Takara Bio Inc. (Otsu, Japan) and Stratagene (La Jolla, CA, USA). All chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA), Wako Pure Chemicals Inc (Osaka, Japan).

Construction of plasmids for wild-type BmGlcDH-III, -IV and its mutants

To express recombinant BmGlcDH-III and BmGlcDH-IV with N-terminal His-tags, the corresponding genes (gene bank accession number D10625 and D10626, respectively) were cloned into the *Nde*I and *Xho*I sites of vector pET28a (Novagen). *Bacillus megaterium* IAM1030 has four GlcDH isozymes (BmGlcDH-I–IV), and the nucleotide sequences of *gdhIII* and *gdhIV* are almost identical. Therefore, the upstream and downstream regions of *gdhIV* were initially amplified using PCR with *Pfu* Turbo DNA polymerase and the specific primers. The *B. megaterium* IAM1030 genome DNA was used as the template. Finally, *gdhIII* and *gdhIV* were amplified using PCR with *Pfu* Turbo DNA polymerase and the specific primers. BmGlcDH-IV single mutants were prepared using inverse PCR with appropriate synthetic oligonucleotides as primers, and pET28a-*gdhIV* as the template.

Expression and purification of the recombinant BmGlcDH-III, -IV and its mutants

Transformed *Escherichia coli* BL21-CodonPlus(DE3)-RIL cells were grown at 37 °C in 100 mL of LB medium supplemented with 20 $\mu g \cdot mL^{-1}$ kanamycin and 34 $\mu g \cdot mL^{-1}$ chloramphenicol. This culture was then inoculated at a ratio of 100 mL of overnight culture to 900 mL of LB medium. Gene expression was induced by the addition of 0.4 mM isopropyl thio-β-D-galactoside, and the culture was incubated overnight at 30 °C. The cells were harvested from the culture by centrifugation at 3000 ***g*** for 10 min. The resultant cell pellet was then resuspended in buffer A (50 mM sodium phosphate, 300 mM NaCl, pH 8.0) containing 1 $mg \cdot mL^{-1}$ lysozyme and about 50 $U \cdot mL^{-1}$ benzonase. The cell suspension was incubated on ice for 30 min and sonicated for 10 min. After centrifugation at 27 200 ***g*** for 20 min, the supernatant was loaded onto a nickel affinity column (HIS-SELECT resin; Sigma-Aldrich (St. Louis, MO, USA)) equilibrated with buffer A. Bound proteins were washed with buffer B (50 mM sodium phosphate, 300 mM NaCl and 10% glycerol, pH 6.0) and eluted using a linear gradient of 0–0.4 M imidazole in buffer B. The

purified fractions were pooled and dialyzed against 50 mM sodium phosphate, 2 M NaCl and 20% glycerol (pH 6.5) and stored at −20 °C.

Enzyme assay

The activities of wild-type BmGlcDH-III, BmGlcDH-IV and its variants were spectrophotometrically assayed by measuring the rate of reduction of NAD^+ in a 3 mL assay mixture at 340 nm at 37 °C. One unit of activity was defined as a reduction of 1 μmol of NADH per minute. Protein concentrations were determined using the Bradford assay with BSA as a standard [35]. The assay mixture contained 148 mM of appropriate substrates, 85.3 mM Tris/HCl pH 8.0 or 85.3 mM potassium phosphate pH 6.0, and 3.66 mM NAD^+ for activity measurements. To examine the pH dependence of the enzyme activity, 85.3 mM potassium phosphate buffer (pH range from 6.0 to 7.5) or 85.3 mM Tris/HCl buffer (pH range from 8.0 to 9.0) was used for the assay. The K_M and V_{max} values for D-glucose were estimated from Michaelis–Menten kinetics after triplicate experiments with varying D-glucose concentrations (2.4–1400 mM; for details see Table 2). To examine the thermostability of wild-type BmGlcDH-III, BmGlcDH-IV and its variants, the remaining activity for D-glucose was determined after incubation for 0, 20 and 40 min at 65 °C. The assay mixture for the thermostability examination contained 0.1 $mg \cdot mL^{-1}$ enzyme in 50 mM sodium phosphate pH 8.0 and 2.0 M NaCl.

Crystallization experiment

For the crystallization experiments, protein concentration was adjusted to 10.0 $mg \cdot mL^{-1}$ in 20 mM Bistris pH 6.5 and 200 mM NaCl. For the crystallization of the cofactor complex, 4 mM NADH was added to the sample followed by an overnight incubation at 4 °C. Crystallization screens were performed using the commercially available sparse-matrix screens (Hampton Research (Aliso Viejo, CA, USA) and Emerald BioStructure (Bainbridge Island, WA, USA)), and diffraction-quality crystals suitable for X-ray structure analysis were obtained by optimizing the buffer pH and precipitant concentration. BmGlcDH-IV in ligand-free form and in complex with NADH was crystallized using a reservoir solution containing 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium chloride and 50% poly(ethylene glycol) 200 by using the hanging-drop vapor diffusion method. BmGlcDH-IV in complex with D-glucose was crystallized using a reservoir solution containing 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium chloride, 51% poly(ethylene glycol) 200 and 400 mM D-glucose by the hanging-drop vapor diffusion method at 293 K. G259A mutant crystals were obtained using a micro-seeding technique in a drop with reservoir solution containing 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium chloride, 50.5% poly(ethylene glycol) 200 and 400 mM D-glucose. Crystals of A258F mutant were obtained using the hanging-drop vapor diffusion method with reservoir solution containing 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium chloride, 50% poly (ethylene glycol) 200 and 400 mM D-glucose. For crystallization of these mutants, 4 mM NADH was added to the sample, followed by overnight incubation at 4 °C.

X-ray diffraction studies and structure determination

The X-ray diffraction data were collected using a charge-coupled device detector (ADSC) under cryogenic conditions (100 K) at beamline AR-NW12A, AR-NE3A or BL-5A, Photon Factory (Tsukuba, Japan). These data were indexed, integrated and merged using the HKL2000 program package [36]. Crystals of wild-type BmGlcDH-IV and G259A mutant belong to the monoclinic $C2$ space group with unit-cell dimensions a = 65 Å, b = 129 Å, c = 64 Å and β = 111°. The A258F mutant was crystallized in different crystal form: the space group $P2_1$ with unit-cell parameters a = 57 Å, b = 127 Å, c = 89 Å and β = 95°. The structure of BmGlcDH-IV was solved via the molecular replacement method by using the program MOLREP [37] and by applying the structure of BmGlcDH-IWG3 (PDB code 1GCO) [27] as a search model. Atomic coordinates and atomic displacement parameters were refined with REFMAC5 program [38], and a manual model fitting was performed using the graphic program COOT [39]. The stereochemical quality of the final refined model was verified using the program PROCHECK [40]. Data collection and refinement statistics are provided in Table 1. The atomic coordinates and structure factor amplitudes for BmGlcDH-IV in ligand-free form, in complex with NADH, in complex with D-glucose, G259A mutant in ligand-free form, and A258F mutant in complex with D-glucose and NADH were deposited in the RCSB Protein Data Bank (http://www.rcsb.org) under the accession codes 3AUS, 3AUT, 3AUU, 3AY6 and 3AY7, respectively. Molecular illustrations were prepared using the program PYMOL (DeLano Scientific; http://pymol.sourceforge.net/).

Acknowledgements

We thank beamline scientists at Photon Factory (Tsukuba, Japan) for their kind assistance with X-ray diffraction experiments. We also thank N. Tamura for technical support. This work was supported by a grant from the National Institute of Advanced Industrial Science and Technology (AIST), Japan.

Conflict of interest

The authors declare no conflict of interest.

References

1 Chaparro-Riggers JF, Rogers TA, Vázquez-Figueroa E & Bommarius AS (2007) Comparison of three enoate reductases and their potential use for biotransformations. *Adv Synth Catal* **349**, 1521–1531.

2 Kataoka M, Sri Rohani LP, Wada M, Kita K, Yanase H, Urabe I & Shimizu S (1998) *Escherichia coli* transformant expressing the glucose dehydrogenase gene from *Bacillus megaterium* as a cofactor regenerator in a chiral alcohol production system. *Biosci Biotechnol Biochem* **62**, 167–169.

3 Imoto N, Nishioka T & Tamura T (2011) Permeabilization induced by lipid II-targeting lantibiotic nisin and its effect on the bioconversion of vitamin D_3 to 25-hydroxyvitamin D_3 by *Rhodococcus erythropolis*. *Biochem Biophys Res Commun* **405**, 393–398.

4 Dokter P, Frank J & Duine JA (1986) Purification and characterization of quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus* L.M.D. 79.41. *Biochem J* **239**, 163–167.

5 Igarashi S, Ohtera T, Yoshida H, Witarto AB & Sode K (1999) Construction and characterization of mutant water-soluble PQQ glucose dehydrogenases with altered K_M values: site-directed mutagenesis studies on the putative active site. *Biochem Biophys Res Commun* **264**, 820–824.

6 Igarashi S, Okuda J, Ikebukuro K & Sode K (2004) Molecular engineering of PQQGDH and its applications. *Arch Biochem Biophys* **428**, 52–63.

7 Hamamatsu N, Suzumura A, Nomiya Y, Sato M, Aita T, Nakajima M, Husimi Y & Shibanaka Y (2006) Modified substrate specificity of pyrroloquinoline quinone glucose dehydrogenase by biased mutation assembling with optimized amino acid substitution. *Appl Microbiol Biotechnol* **73**, 607–617.

8 Bak T (1967) Studies on glucose dehydrogenase of *aspergillus oryzae*: III. General enzymatic properties. *Biochem Biophys Acta* **146**, 317–327.

9 Tsujimura S, Kojima S, Kano K, Ikeda T, Sato M, Sanada H & Omura H (2006) Novel FAD-dependent glucose dehydrogenase for a dioxygen-insensitive glucose biosensor. *Biosci Biotechnol Biochem* **70**, 654–659.

10 Yamaoka H, Yamashita Y, Ferri S & Sode K (2008) Site directed mutagenesis studies of FAD-dependent glucose dehydrogenase catalytic subunit of *Burkholderia cepacia*. *Biotechnol Lett* **3**, 1967–1972.

11 Pauly HE & Pfleiderer G (1975) D-Glucose dehydrogenase from *Bacillus megaterium* M 1286: purification, properties and structure. *Hoppe Seylers Z Physiol Chem* **356**, 1613–1623.

12 Mitamura T, Urabe I & Okada H (1989) Enzymatic properties of isozymes and variants of glucose dehydrogenase from *Bacillus megaterium*. *Eur J Biochem* **186**, 389–393.

13 Nagao T, Mitamura T, Wang XH, Negoro S, Yomo T, Urabe I & Okada H (1992) Cloning, nucleotide sequences, and enzymatic properties of glucose dehydrogenase isozymes from *Bacillus megaterium* IAM1030. *J Bacteriol* **174**, 5013–5020.

14 Jörnvall H, Persson B, Krook M, Atrian S, Gonzàlez-Duarte R, Jeffery J & Ghosh D (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **34**, 6003–6013.

15 Pauly HE & Pfleiderer G (1977) Conformation and functional aspects of the reversible dissociation and denaturation of glucose dehydrogenase. *Biochemistry* **16**, 4599–4604.

16 Maurer E & Pfleiderer G (1985) Reversible pH-induced dissociation of glucose dehydrogenase from *Bacillus megaterium* I. Conformational and functional changes. *Biochim Biophys Acta* **827**, 381–388.

17 Maurer E & Pfleiderer G (1987) Reversible pH-induced dissociation of glucose dehydrogenase from *Bacillus megaterium* II. Kinetics and mechanism. *Z Naturforsch* **42c**, 905–915.

18 Vázquez-Figueroa E, Chaparro-Riggers J & Bommarius AS (2007) Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. *ChemBioChem* **8**, 2295–2301.

19 Makino K, Negoro S, Urabe I & Okada H (1989) Stability-increasing mutants of glucose dehydrogenase from *Bacillus megaterium* IWG3 . *J Biol Chem* **264**, 6381–6385.

20 Nagao T, Makino Y, Yamamoto K, Urabe I & Okada H (1989) Stability-increasing mutants of glucose dehydrogenase. *FEBS Lett* **253**, 113–116.

21 Baik SH, Ide T, Yoshida Y, Kagami O & Harayama S (2003) Significantly enhanced stability of glucose dehydrogenase IWG3 from *Bacillus megaterium* by directed evolution. *Appl Microbiol Biotechnol* **61**, 329–335.

22 Baik SH, Michel F, Aghajari N, Haser R & Harayama S (2005) Cooperative effect of two surface amino acid mutations (Q252L and E170K) in glucose dehydrogenase from *Bacillus megaterium* IWG3 on stabilization of its oligomeric state. *Appl Environ Microbiol* **71**, 3285–3293.

23 Yasutake Y, Nishiya Y, Tamura N & Tamura T (2007) Structural insights into unique substrate selectivity of *Thermoplasma acidophilum* D-aldohexose dehydrogenase. *J Mol Biol* **367**, 1034–1046.

24 Kubota K, Nagata K, Okai M, Miyazono K, Soemphol W, Ohtsuka J, Yamamura A, Saichana N, Toyama H, Matsushita K *et al.* (2011) The crystal structure of L-sorbose reductase from *Gluconobacter frateurii* complexed with NADPH and L-sorbose. *J Mol Biol* **407**, 543–555.

25 Tanaka N, Nonaka T, Tanabe T, Yoshimoto T, Tsuru D & Mitsui Y (1996) Crystal structures of the binary and ternary complexes of 7α-hydroxysteroid dehydrogenase from *Escherichia coli*. *Biochemistry* **35**, 7715–7730.
26 Benach J, Atrian S, Gonzàlez-Duarte R & Ladenstein R (1999) The catalytic reaction and inhibition mechanism of *Drosophila* alcohol dehydrogenase: observation of an enzyme-bound NAD-ketone adduct at 1.4 Å resolution by X-ray crystallography. *J Mol Biol* **289**, 335–355.
27 Yamamoto K, Kurisu G, Kusunoki M, Tabata S, Urabe I & Osaki S (2001) Crystal structure of glucose dehydrogenase from *Bacillus megaterium* IWG3 at 1.7 Å resolution. *J Biochem* **129**, 303–312.
28 Bottoms CA, Smith PE & Tanner JJ (2002) A structurally conserved water molecule in Rossmann dinucleotide-binding domains. *Protein Sci* **11**, 2125–2137.
29 Duax WL, Pletnev V, Addlagatta A, Bruenn J & Weeks CM (2003) Rational proteomics. I. Fingerprint identification and cofactor specificity in the short-chain oxidoreductase (SCOR) enzyme family. *Proteins* **53**, 931–943.
30 Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H & Oppermann U (2002) Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J Biol Chem* **277**, 25677–25684.
31 Ghosh D, Sawicki M, Pletnev V, Erman M, Ohno S, Nakajin S & Duax WL (2001) Porcine carbonyl reductase: structural basis for a functional monomer in short chain dehydrogenases/reductases. *J Biol Chem* **276**, 18457–18463.
32 Albalat R, Valls M, Fibla J, Atrian S & Gonzàlez-Duarte R (1995) Involvement of the C-terminal tail in the activity of *Drosophila* alcohol dehydrogenase: evaluation of truncated proteins constructed by site-directed mutagenesis. *Eur J Biochem* **233**, 498–505.
33 Nishioka T, Yasutake Y, Nishiya Y, Tamura N & Tamura T (2009) C-terminal tail derived from the neighboring subunit is critical for the activity of *Thermoplasma acidophilum* D-aldohexose dehydrogenase. *Proteins* **74**, 801–807.
34 Parkkinen T, Boer H, Jänis J, Andberg M, Penttilä M, Koivula A & Rouvinen J (2011) Crystal structure of uronate dehydrogenase from *Agrobacterium tumefaciens*. *J Biol Chem* **286**, 27294–27300.
35 Bradford MM (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254.
36 Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307–326.
37 Vagin A & Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30**, 1022–1025.
38 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr Sect D Biol Crystallogr* **53**, 240–255.
39 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* **60**, 2126–2132.
40 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.

Supporting information

The following supplementary information is available:
Fig. S1. Stereo-view superimposition of the cofactor binding site structures of BmGlcDH-IV and BmGlcDH-IWG3.
Fig. S2. Stereo-view $F_o - F_c$ omit map for the bound β-D-glucose in the wild-type BmGlcDH-IV contoured at 3.0σ level.
Table S1. Hydrogen bonding interactions between BmGlcDH-IV and D-glucose.

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 reorganized 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.

Food Research International 49 (2012) 567–571

Contents lists available at SciVerse ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Properties of emu (*Dromaius novaehollandiae*) albumen proteins

Junko Takeuchi [a,*], Kenji Maehashi [b], Yoshiaki Yasutake [c], Yoshiki Muramatsu [d], Keita Miyata [d], Toshihiro Watanabe [d], Toshio Nagashima [d]

[a] *Bio-resources Institute, Faculty of Bio-industry, Tokyo University of Agriculture, Japan*
[b] *Department of Fermentation Science, Faculty of Applied Bio-Science, Tokyo University of Agriculture, Japan*
[c] *Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Japan*
[d] *Department of Food and Cosmetic Science, Faculty of Bio-industry, Tokyo University of Agriculture, Japan*

ARTICLE INFO

Article history:
Received 14 May 2012
Accepted 21 July 2012

Keywords:
Emu
Egg
Albumen
Ovalbumin
Ovotransferrin
Tenp

ABSTRACT

The emu, a flightless bird native to Australia, lays eggs that have different properties than hen eggs: emu eggs begin to coagulate at a lower temperature than do hen eggs, and hard-boiled emu albumen has less gel strength than hen albumen. To investigate the mechanisms behind the differences, three major albumen proteins, ovotransferrin, ovalbumin, and tenp, were analyzed calorimetrically and physicochemically. In the calorimetrical analysis, emu ovotransferrin and ovalbumin exhibited similar thermostability to their hen counterparts. Tenp, a major component of emu but not hen albumen, showed high thermostability. Ovotransferrin, the major protein in emu albumen, produced a softer gel than ovalbumin when denatured in boiling water. Tenp did not coagulate completely in boiling water. The above results suggest the usefulness of emu eggs for processed food production. Changes in the antigenicity of emu ovalbumin during digestion were also examined in vitro.

1. Introduction

The emu (*Dromaius novaehollandiae*) is a flightless bird from Australia that produces eggs only during the winter. Emu eggs weigh 10 times more than hen eggs and are known to possess properties amenable to gelation, whipping, and foaming (Takeuchi & Nagashima, 2010). For example, emu eggs begin to coagulate at a lower temperature than do hen eggs. Hard-boiled emu albumen has 1/3 the gel strength of hen albumen. Foam stability is also superior in the emu egg, which means, for example, that a cake made with emu eggs is softer than one made with hen egg products. Local confectioneries are utilizing emu eggs because these unique properties are advantageous for various production processes.

Emu albumen is composed of components similar to those of hen albumen but at a different ratio (Maehashi et al., 2010). The difference in albumen quality could be ascribed to this ratio. However, the properties of each composing protein have not yet been analyzed precisely. Thus, this study investigated the processing properties of the major constituent proteins. Emu albumen contains the following major proteins: ovotransferrin, ovalbumin, and tenp protein (abbreviated as tenp). The ovotransferrin in emu albumen is 75.3% identical to the hen counterpart (Maehashi et al., 2010). It is the most abundant protein component of an emu egg but the second most abundant protein component of a hen egg. Therefore, it is conceivable that ovotransferrin strongly influences the properties of emu eggs. Hen ovotransferrin is known to coagulate at the low temperature of 65 °C (Donovan, Mapes, Davis, & Garibaldi, 1997). It is an ion-binding protein that is known to stabilize by ion loading at the binding sites (Kurokawa, Mikami, & Hirose, 1995; Lin, Mason, Woodworth, & Brandts, 1991; Mine, 1995). However, little is known about the properties of emu ovotransferrin.

Emu ovalbumin behaves like a 102-kDa high-molecular-weight protein on SDS-PAGE analysis, although its primary structure is 71.3% identical to 44.5-kDa hen ovalbumin. Interestingly, it has been reported that emu ovalbumin is not recognized by anti-hen ovalbumin polyclonal antibody (Maehashi et al., 2010). However, consumers may hesitate to use emu egg for allergic patients, because of the high level of similarity between the primary structures of emu and hen ovalbumin. To confirm the safety of emu eggs, it must be proven that no antigenic fragments emerge from emu ovalbumin during digestion. About the properties of emu ovalbumin, the mobility pattern and difference in antigen recognition suggest ovalbumin polymer formation occurs in addition to protein modifications such as phosphorylation and glycosylation. Therefore, it may be possible that there are some differences in the properties of hen and emu ovalbumin.

Tenp has been isolated as a minor component of infertile hen albumen (Guérin-Dubiard et al., 2006). It is also known to be a putative membrane protein from which mRNA is transiently expressed during embryonic development (Yan & Wang, 1998). In the emu egg, tenp constitutes 16% of the total albumen protein; thus, its characteristics may play a role in the unique properties of the emu egg.

In this study, we separately purified ovotransferrin, ovalbumin, and tenp and then performed a calorimetric analysis and heat denaturing

* Corresponding author at: 196 Yasaka, Abashiri, Hokkaido, Japan, 099-2493. Tel.: +81 152 48 3800; fax: +81 152 48 3839.
E-mail address: j3takeuc@bioindustry.nodai.ac.jp (J. Takeuchi).

0963-9969/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodres.2012.07.045

gel strength test. Ovalbumin was also subjected to immunoblotting using anti-hen ovalbumin antibody to examine antigenicity during in vitro digestion.

2. Materials and methods

2.1. Materials

Commercially available fertile emu eggs were kindly provided by Tokyo Nodai Bio-industry Co., LTD. (Hokkaido, Japan). Fertile hen eggs (White leghorn) were provided by Dr. Gaku Shimoi of Tokyo University of Agriculture. All experimental reagents were purchased from Kanto Chemical Co., INC. (Tokyo, Japan) unless otherwise noted.

2.2. Purification of emu albumen components

Fresh emu albumen was removed from the yolk and homogenized in a whirling blender (AM-7, Nihonseiki, Tokyo, Japan). The pH of the homogenized albumen was adjusted to 6.0 by adding HCl, and the albumen was stirred overnight at 4 °C to produce ovomucin precipitate (Guerin-Dubiard et al., 2005). The precipitated ovomucin was removed by centrifugation for 10 min at 17,000 *g* (Avanti HP-30I, Beckman Coulter, CA, USA). The supernatant was dialyzed against 50 mM Tris–HCl (pH 8.0) and 50 mM NaCl and then stored at −20 °C until use. This mucin-free albumen was loaded onto an anion-exchanging column (DEAE 650M, Tosoh, Tokyo, Japan) that had been equilibrated with 50 mM Tris–HCl (pH 8.0) and 50 mM NaCl. Protein elution was performed by gradually increasing the NaCl concentration to 500 mM. Ovotransferrin (78 kDa) was collected in an unbound fraction. Ovalbumin (102 kDa) and tenp (47 kDa) were jointly eluted out as the NaCl concentration increased. The eluate containing ovalbumin and tenp was dialyzed and then loaded onto a hydroxyapatite Biogel HTP column (BioRad, Hercules, CA, USA) that had been equilibrated with 10 mM potassium phosphate (pH 6.7) and 50 mM NaCl. Ovalbumin and tenp were separately eluted out as the phosphate concentration increased. The fractions that contained both ovalbumin and tenp were reloaded onto the hydroxyapatite Biogel HTP to clearly fractionate the two proteins. Isolated 102-kDa and 47-kDa proteins were analyzed by an amino acid sequencer and identified as ovalbumin and tenp, respectively (data not shown). Purified protein was dialyzed against PBS (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 140 mM NaCl, and 2.7 mM KCl, pH 6.8), and then concentrated using a membrane filter (10-kDa pore; Millipore, Billerica, MA, USA).

The crude mixture without ovotransferrin that was to be used for the calorimetric experiment was prepared by anion exchange chromatography of the mucin-free emu albumen as described above. Elution was performed isocratically with 0.5 M NaCl. The crude mixture without tenp was prepared by overloading mucin-free emu albumen into the DEAE 650M column and collecting the flow-through fraction.

The recombinant ovalbumin, the hexahistidine (6His)-tagged version at the carboxy terminus, was purified through Ni-NTA superflow resin (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The protein was eluted by increasing the imidazole concentration, collected, and subjected to further purification by hydroxyapatite Biogel HTP, as described above.

2.3. Calorimetric analysis

The purified protein in PBS was concentrated using a membrane filter to obtain a protein concentration of 5~10 mg/ml. Protein solution (60 μl) was transferred into an aluminum canister (AL-70 capsule, Epolead Service Inc., Chiba, Japan) that had been oxidized, and then sealed by an automatic sealer (Epolead Service, Inc.). The canister was set into a differential scanning calorimeter (DSC-6100, Seiko Instruments Inc., Chiba, Japan) along with a control canister that contained PBS. Calorimetric analysis was performed by raising the temperature from 30 °C to 150 °C at a rate of 1 °C/min.

2.4. Gel strength analysis

Two hundred microliters of the protein solution was transferred to a 2-ml plastic tube and boiled for 10 min. The tube was immediately cooled under running water. After cooling, the tube was fixed on a stand and then analyzed using a Tensipresser (TTP-50BXII-2006, Taketomo Electric, Tokyo, Japan), which was equipped with a cylindrical plunger whose diameter was 3 mm. The test was run in one-bite mode with a bite speed of 1 mm/sec. Experiments were performed in triplicates. Resulting mean values were statistically evaluated by Welch's *t*-test.

2.5. In vitro digestion and immunoblotting

The purified emu ovalbumin was digested with simulated gastric fluid (SGF) as previously reported (Astwood, Leach, & Fuchs, 1996). The SGF contained 0.02%(w/v) pepsin (from porcine gastric mucosa, Sigma-Aldrich, St Louis, MO, USA) and 30 mM of NaCl. The SGF was adjusted to a pH of 1.2 using HCl. The reaction with SGF was performed as follows: an 80-μl aliquot of SGF and 70 μl of 0.5 mg/ml protein solution were separately preheated at 37 °C. The SGF was then added to the protein solution and continuously incubated. At every time point, 40 μl of 160 mM Na_2CO_3 was added to the reaction mixture to quench the reaction. A 10-μl reaction mixture was mixed with 2× Laemmli SDS-PAGE buffer, boiled and subjected to SDS-PAGE. Hen ovalbumin (Nacalai Tesque, Kyoto, Japan) was used as a control. For the denaturing condition, the ovalbumin protein solution was boiled for 10 min prior to digestion in SGF.

The gels were electroblotted onto PVDF membranes (GE Healthcare UK Ltd., Little Chalfont, England). The membranes were then blocked by incubation in 5% non-fat dried milk in TBST (25 mM Tris–HCl (pH 7.5) and 15 mM NaCl, 0.1% Tween-20) for 1 h at room temperature. A rabbit anti-hen ovalbumin polyclonal antibody (Millipore) was diluted to 1:8000 in TBST and used as a primary antibody. A secondary antibody (HRP conjugated goat anti-rabbit IgG (GE Healthcare)) was diluted to 1:10,000 in TBST. Detection was performed according to the manufacturer's instruction, and the illuminant was captured by a light capture II AE-6981 image scanner (ATTO Corp., Tokyo, Japan).

3. Results and discussion

3.1. Denaturing temperature of emu albumen proteins

Because emu egg behaves differently than hen egg during heat denaturation, it was hypothesized that the albumen proteins play important roles. First, the thermostability of crude albumen during a temperature shift was examined to confirm the previously reported results (Nakamura et al., 2011). As shown in Fig. 1A, three peaks were observed in the emu albumen. The exact peak temperatures are summarized in Table 1. Two obvious peaks were detected in a lower temperature range in the emu albumen and corresponded to the hen albumen. The peak observed above 90 °C was small but unique in the emu albumen. To identify which component was responsible for which peak, we then fractionated the emu albumen. The prepared fraction was analyzed by SDS-PAGE prior to the calorimetric analysis. The tenp-removed fraction is shown in Fig. 2A. This fraction did not show a band at 45 kDa, and apparently lost the peak above 90 °C (Fig. 1B) even though it had unidentified peaks in the lower temperature range. The ovotransferrin-removed fraction was smeared in an SDS-PAGE analysis (Fig. 2B) but showed only two obvious peaks at 74.2 °C and 91.9 °C (Fig. 1B). Lastly, we separately purified and then examined the three major components of emu albumen: ovotransferrin, ovalbumin, and tenp. The results of the SDS-PAGE analysis of the purified proteins are shown in Fig. 2C.

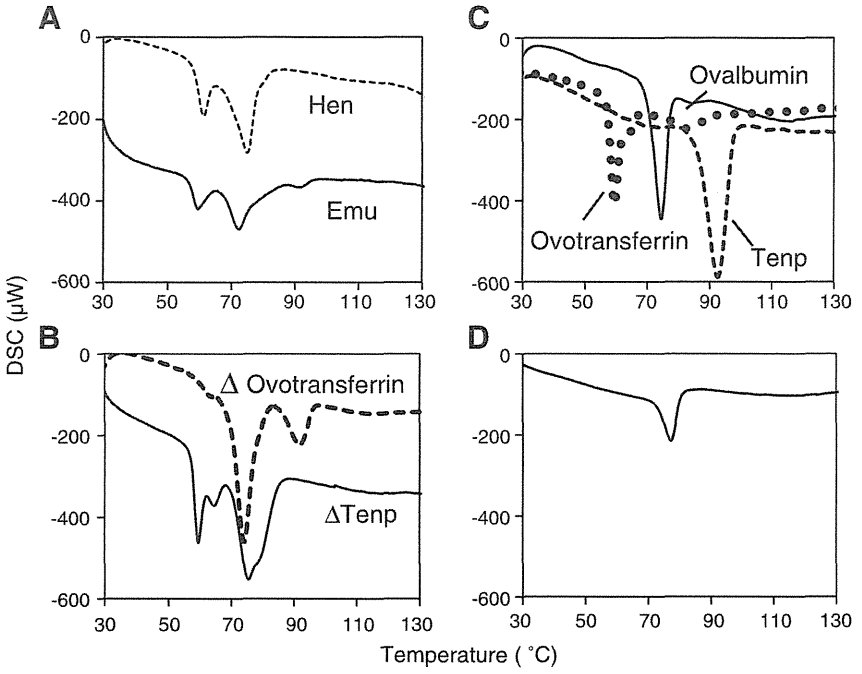

Fig. 1. DSC analysis of crude and purified proteins. (A) Emu and hen albumen were directly analyzed as described in materials and methods. (B) Either tenp-removed or ovotransferrin-removed emu albumen fractions were analyzed. (C) Each of the purified proteins from the emu albumen was analyzed. (D) *Escherichia coli*-expressed recombinant ovalbumin was analyzed.

Each protein exhibited a very clear peak in the calorimetric analysis (Fig. 1C). Ovotransferrin had a peak at 59.5 °C, ovalbumin had a peak at 77.3 °C, and tenp had a peak at 92.9 °C; these results correspond with those shown in Fig. 1B. Thus, three major peaks in crude albumen were identified and the denaturing temperatures of emu ovalbumin and tenp were determined for the first time. It has been reported that emu ovotransferrin showed two peaks at 60.0 °C and 80.0 °C (Nakamura et al., 2011). In our experiment, there was also a signal at approximately 80 °C, but it was broad and relatively small (Fig. 1C). The ion-saturated form of hen ovotransferrin is known to be stable and denatures at around 80 °C (Lin et al., 1991), in contrast to the main denaturation peak at 57.3 °C (Johnson & Zabik, 1981); therefore, the broad peak at approximately 80 °C in our experiment was probably from an ion-bound form. During the concentration of purified ovotransferrin, the sample appeared pale red, which suggests ferric ion binding in its native state.

Emu egg ovalbumin is detected as a 102-kDa band in SDS-PAGE analysis. We expected to see a pattern different from that of the hen counterpart in the calorimetric analysis if it formed a tight dimer. However, purified ovalbumin showed only a single peak at 77.3 °C when it was heated to 150 °C, which is similar to the denaturation pattern of hen egg (Arii, Takahashi, Tatsumi, & Hirose, 1999; Yamasaki, Takahashi, & Hirose, 2003). To confirm the result, we prepared recombinant ovalbumin, which showed two bands of 43 kDa and 77 kDa in an SDS-PAGE analysis (Fig. 2D). As a result, the recombinant ovalbumin showed one clear peak at 74.7 °C (Fig. 1D). Thus, it was determined that the different mobility of ovalbumin in SDS-PAGE does not lead to any changes in its denaturing property. The denaturing temperature of purified emu ovalbumin was slightly higher than that of either the recombinant ovalbumin or the crude fractions. Such a difference may have occurred owing to the carboxy-terminus 6His-tag or to chance. Thermostability differences between native and recombinant hen ovalbumin have also been previously observed (Arii et al., 1999). Amino acid sequences are 75.3% similar between hen and emu ovalbumin (Maehashi et al., 2010). Similarity in denaturing temperature between hen and emu ovalbumin suggests that critical amino acids for denaturation might be conserved. To determine the region that causes this mobility difference on SDS-PAGE, mutational analysis in either hen or emu ovalbumin is ongoing.

Table 1
Peak temperature in DSC analysis.

Sample		Peak temperature (°C)		
Crude albumen	Emu	60.2	72.9	92.3
	Hen	61.4	75.0	-
Fraction	ΔTenp	59.6	75.5	-
	ΔOvotransferrin	-	74.2	91.9
Ovalbumin		-	77.3	-
Ovalbumin-6His		-	74.7	-
Ovotransferrin		59.5	-	-
Tenp		-	-	92.9

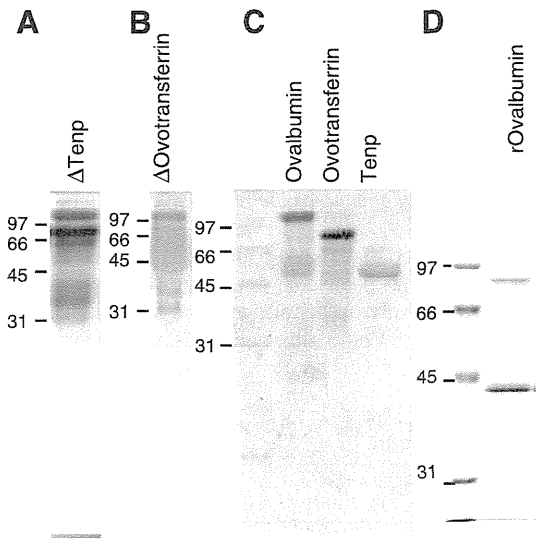

Fig. 2. Fractionated proteins. The tenp-removed fraction (A) and the ovotransferrin-removed fraction (B) were analyzed via SDS-PAGE. (C) The purified ovalbumin, ovotransferrin, and tenp proteins. (D) Purified recombinant ovalbumin-6His protein.

3.2. Properties of heat denatured gel

Next, the strength of heat-denatured gel was examined. Because of the limited sample amount, a micro-scale experiment was established. To determine the reproducibility and accuracy of this micro-scale experiment, crude albumen (approximate protein content is 9% in emu egg and 10.5% in hen egg) was first examined. A sharp peak appeared when the gel broke as the plunger moved down. The pressure at the peak was designated as the gel strength. As shown in Table 2, the results were moderately reproduced in triplicate experiments. The standard deviations were reasonably small, which convinced us that this micro-scale experiment was sufficiently accurate. The results showed that the hen albumen gel was significantly stronger than the emu albumen gel, which corresponds to a previous report (Takeuchi & Nagashima, 2010). The plunger insertion distance before gel breakage represents elasticity, although the plunger shape may influence the results. This experiment, in which we used a cylindrical plunger, revealed that the emu albumen gel was less elastic than the hen albumen gel.

Next, each purified protein was heat-denatured and subjected to analysis. Ovalbumin and ovotransferrin showed obvious peaks when the gel broke; however, there were no apparent peaks in the tenp gel analysis. The plunger proceeded as far as 9.99 mm in the tenp gel with very little pressure (Table 3). Tenp was heat-denatured but only formed viscous paste. In contrast, ovalbumin formed a firm gel, which had significantly higher strength than other proteins. The elasticity of the ovalbumin gel was also high, which implies that breaking ovalbumin gel requires considerably large amount of energy. The ovotransferrin gel had less than half the strength of the ovalbumin gel, and its elasticity was significantly less than that of ovalbumin gel. Therefore, it was revealed that ovotransferrin is responsible for the soft property of emu albumen. The amount of tenp in emu albumen must contribute to the unique properties of the albumen. Specifically, the nature of the tenp gel explains why bakery products made from emu eggs, such as cakes and crème brûlée, are often perceived as undercooked. In case of hen egg albumen, it has been reported that the ovotransferrin gel had about 2/3 of the strength of ovalbumin gel (Johnson & Zabik, 1981). It cannot be directly compared with our results, although, it suggests that the content of two major components influences the physical property of albumen gel.

To this end, ovotransferrin, ovalbumin, and tenp were combined at a ratio of 50:25:25 to roughly reconstitute crude emu albumen. This mixture (10% protein concentration) was heat denatured and subjected to a gel strength test. As shown in Table 3, the gel was firmer than the crude emu albumen and thus did not completely reproduce the gelation property of crude emu albumen. The above result suggests that additional factors influence the fragility of emu albumen gel. The interaction between major and minor components may influence the fragility of emu albumen because the formation of intermolecular bonds by heat is important for the transition of egg albumen to gel (Mine, 1995).

3.3. Antigenic property of emu ovalbumin

It has been reported that 102-kDa emu ovalbumin is not recognized by anti-hen ovalbumin polyclonal antibody (Maehashi et al., 2010). However, it is not clear whether antigenic peptides are released during human digestion of emu albumen. Thus, we performed in vitro digestion with simulated gastric fluid. First, hen ovalbumin was used as a control. During the raw protein was directly digested, the 45-kDa band was stable in SDS-PAGE analysis (Fig. 3A), even though the anti-hen ovalbumin-reactive protein gradually decreased in the immunoblotting (Fig. 3B). Conversely, it disappeared in 15 s in the SDS-PAGE analysis when the heat-denatured hen ovalbumin was used for the experiment. In immunoblotting, a ladder of smaller bands was detected up to 60 s. When the raw emu albumen was digested, a slight disappearance of the 102-kDa emu ovalbumin was observed in a time-dependent manner, as shown in Fig. 3C. In contrast, minor bands at approximately 45 kDa accumulated in a time-dependent manner. Immunoblotting indicated that the 102-kDa band was reactive to anti-hen ovalbumin. Another band at approximately 45 kDa was significantly reactive to the anti-hen ovalbumin antibody. It was visible from time 0 and accumulated in a time-dependent manner (Fig. 3D). Although the manufacturer of the polyclonal antibody was different from that in a previous report (Maehashi et al., 2010), the recognition of emu ovalbumin by an anti-hen ovalbumin antibody was clearly demonstrated. It was hypothesized that the structure of emu ovalbumin would deteriorate to 45 kDa, which would make the 45-kDa band more recognizable to the antibody. When the heat-denatured emu ovalbumin was used for the experiment, the 102-kDa band completely disappeared in 15 s, and a vague band at approximately 30-kDa appeared at 15 s and then disappeared by 60 s (Fig. 3C). The 30-kDa band was more readily detected in immunoblotting (Fig. 3D). These results imply that emu ovalbumin is antigenic to anti-hen ovalbumin antibody; the affinity particularly increases when the 102-kDa fragment is converted into smaller fragments. Heat denaturation accelerates digestion; however, it produces antigenic peptides during digestion. To examine the antigenicity of recombinant protein, the same experiment was performed with purified recombinant emu ovalbumin. Unlike native ovalbumin from either hen or emu, raw recombinant ovalbumin was quickly digested by pepsin (Fig. 3E). In the SDS-PAGE analysis, two possible dimer and monomer bands (85 kDa and 43 kDa) were visible at time 0, but the dimer band disappeared in 15 s. At the same time, a band that was slightly smaller than the monomer appeared, and then all of the bands diminished by 60 s. In immunoblotting, a possible dimer and monomer were evenly and readily detected at time 0; the zone between the two bands was smeared (Fig. 3F). During the raw protein was digested, an intense band that was slightly smaller than the monomer appeared at the position at 15 s. Additionally, more bands appeared around the 25-kDa position. When heat denatured recombinant ovalbumin was used for the experiment, both possible dimer and monomer bands were quickly digested and diminished in 15 s (Fig. 3E), which left fragments of approximately 15 kDa. Immunoblotting revealed that anti-hen ovalbumin antibody-reactive intermediates of approximately 30 kDa and 15 kDa appeared during digestion. Bands smaller than 30 kDa were not detectable in the SDS-PAGE analysis, which suggests that those fragments were highly reactive to the anti-hen ovalbumin antibody.

In the above experiments, the apparent cross reactivity of emu ovalbumin against anti-hen ovalbumin antibody was confirmed. Although it had been anticipated that emu albumen had lower

Table 2
Strength of crude albumen gel.

	Peak pressure ($\times 10^3$ N/m^2)	Distance before breakage (mm)
Hen	148 ± 3^{a}	1.70 ± 0.07^{b}
Emu	3.30 ± 0.56^{a}	0.48 ± 0.10^{b}

Mean values of triplicate experiments ± standard deviation are shown. Values with the same superscript express significant difference ($p<0.01$).

Table 3
Strength of each protein gel.

	Peak pressure ($\times 10^3$ N/m^2)	Distance before breakage (mm)
Ovalbumin	7.25 ± 0.65^{ab}	3.28 ± 0.24^{d}
Ovotransferrin	2.63 ± 0.40^{ac}	0.87 ± 0.17^{d}
Tenp	0.47 ± 0.11^{bc}	>9.99 *
Reconstituted emu albumen	6.28 ± 0.12	0.83 ± 0.08

Mean values of triplicate experiments ± standard deviation are shown.
Values with the same superscript express significant difference (a, b, d: $p<0.01$, c: $p<0.05$).
* The gel did not show apparent breakage under the examined condition.

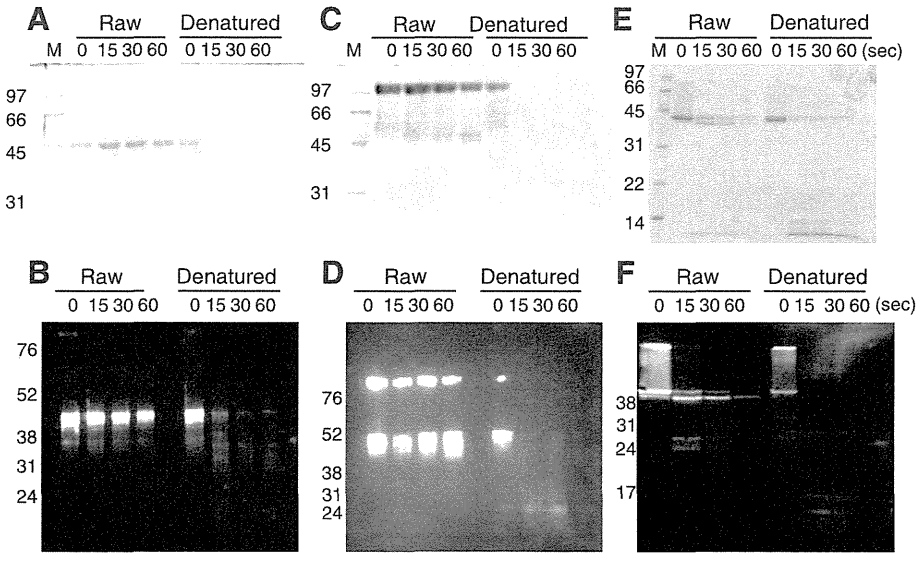

Fig. 3. In vitro digestion of ovalbumin and recognition of ovalbumin by an anti-hen ovalbumin antibody. Commercial hen albumin (A, B), purified native emu ovalbumin (C, D), and recombinant emu ovalbumin (E, F) were digested in vitro with pepsin. At every time point, the resulting protein was either analyzed by SDS-PAGE (A, C, E) or immunoblotting using anti-hen ovalbumin antibody (B, D, F). The numbers to the left of the gels represent molecular weight (kDa). The experiments were performed using either raw or heat-denatured ovalbumin.

allergenicity in egg allergy patients (Maehashi et al., 2010), the above results contradicted this expectation. The result of in vitro digestion suggested that the recombinant ovalbumin possesses a pepsin-sensitive locus. When the locus is cleaved, ovalbumin becomes slightly smaller than a monomer. Due to this cleavage, the dimer structure relaxes, and the dimer band therefore quickly disappears. Additionally, the ovalbumin becomes more sensitive to pepsin after this nibbling. It is expected that this pepsin-sensitive locus is masked in the native form by a modification such as glycosylation. Structural analysis is on going to clarify the above hypothesis.

4. Conclusion

To elucidate the properties of the emu egg, three major proteins that constitute emu albumen were analyzed calorimetrically and physicochemically. Ovotransferrin, the most abundant component was shown to be responsible for making a soft gel. Emu ovalbumin, which mobilizes differently from its hen counterpart in SDS-PAGE analysis, was shown to have similar denaturing temperature to hen ovalbumin. Tenp, a characteristic constituent of emu albumen, demonstrated high stability to heat. These properties, that explain the unique nature of emu egg, were clearly revealed for the first time in this study. In addition, emu ovalbumin, and its digestive intermediates were shown to cross-react with the anti-hen ovalbumin antibody. This finding suggests that emu egg should not be consumed by patients with egg allergies.

Acknowledgements

We are grateful to Tokyo Nodai Bio-Industry Co. LTD. and Dr. Gaku Shimoi for kindly providing the emu eggs and the hen eggs, respectively. We express special gratitude to Kaoru Shimazaki, Ryoko Kishida, and Junichi Konda for their contributions to the experiments. We thank Dr. Shuki Fujimura for her support with the immunoblotting experiments.

References

Arii, Y., Takahashi, N., Tatsumi, E., & Hirose, M. (1999). Structural properties of recombinant ovalbumin and its transformation into a thermostabilized form by alkaline treatment. *Bioscience, Biotechnology, and Biochemistry, 63*, 1392–1399.

Astwood, J. D., Leach, J. N., & Fuchs, R. L. (1996). Stability of food allergens to digestion in vitro. *Nature Biotechnology, 14*, 1269–1273.

Donovan, J. W., Mapes, C. J., Davis, J. G., & Garibaldi, J. A. (1975). A differential scanning calorimetric study of the stability of egg white to heat denaturation. *Journal of the Science of Food and Agriculture, 26*, 73–83.

Guerin-Dubiard, C., Pasco, M., Hietanen, A., del Bosque, A. Q., Nau, F., & Croguennec, T. (2005). Hen egg white fractionation by ion-exchange chromatography. *Journal of Chromatography. A, 1090*, 58–67.

Guérin-Dubiard, C., Pasco, M., Mollé, D., Désert, C., Croguennec, T., & Nau, F. (2006). Proteomic analysis of hen egg white. *Journal of Agricultural and Food Chemistry, 54*, 3901–3910.

Johnson, T. M., & Zabik, M. E. (1981). Gelation properties of albumen proteins, singly and in combination. *Poultry Science, 60*, 2071–2083.

Kurokawa, H., Mikami, B., & Hirose, M. (1995). Crystal-structure of diferric hen ovotransferrin at 2.4 Angstrom resolution. *Journal of Molecular Biology, 254*, 196–207.

Lin, L. N., Mason, A. B., Woodworth, R. C., & Brandts, J. F. (1991). Calorimetric studies of the binding of ferric ions to ovotransferrin and interactions between binding-sites. *Biochemistry, 30*, 11660–11669.

Maehashi, K., Matano, M., Irisawa, T., Uchino, M., Itagaki, Y., Takano, K., et al. (2010). Primary structure of potential allergenic proteins in emu (*Dromaius novaehollandiae*) egg white. *Journal of Agricultural and Food Chemistry, 58*, 12530–12536.

Mine, Y. (1995). Recent advances in the understanding of egg-white protein functionality. *Trends in Food Science & Technology, 6*, 225–232.

Nakamura, Y., Matsumoto, Y., Irisawa, T., Noguchi, H., Uchino, M., Maehashi, K., et al. (2011). Analysis of emu egg white protein by using two-dimensional polyacrylamide gel electrophoresis. *Food Preservation Science, 37*, 289–293.

Takeuchi, J., & Nagashima, T. (2010). Chemical and physical characterization of *Dromaius novaehollandiae* (emu) eggs. *Food Science and Technology Research, 16*, 149–156.

Yamasaki, M., Takahashi, N., & Hirose, M. (2003). Crystal structure of S-ovalbumin as a non-loop-inserted thermostabilized serpin form. *Journal of Biological Chemistry, 278*, 35524–35530.

Yan, R. T., & Wang, S. Z. (1998). Identification and characterization of tenp, a gene transiently expressed before overt cell differentiation during neurogenesis. *Journal of Neurobiology, 34*, 319–328.

Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*

Kei Kitahara[a], Yoshiaki Yasutake[a], and Kentaro Miyazaki[a,b,1]

[a]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Sapporo 062-8517, Japan; and [b]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Sapporo 062-8517, Japan

Edited by W. Ford Doolittle, Dalhousie University, Halifax, Canada, and approved October 4, 2012 (received for review August 6, 2012)

The bacterial ribosome consists of three rRNA molecules and 57 proteins and plays a crucial role in translating mRNA-encoded information into proteins. Because of the ribosome's structural and mechanistic complexity, it is believed that each ribosomal component coevolves to maintain its function. Unlike 5S rRNA, 16S and 23S rRNAs appear to lack mutational robustness, because they form the structural core of the ribosome. However, using *Escherichia coli* Δ7 (null mutant of operons) as a host, we have recently shown that an active hybrid ribosome whose 16S rRNA has been specifically substituted with that from non–*E. coli* bacteria can be reconstituted in vivo. To investigate the mutational robustness of 16S rRNA and the structural basis for its functionality, we used a metagenomic approach to screen for 16S rRNA genes that complement the growth of *E. coli* Δ7. Various functional genes were obtained from the Gammaproteobacteria and Betaproteobacteria lineages. Despite the large sequence diversity (80.9–99.0% identity with *E. coli* 16S rRNA) of the functional 16S rRNA molecules, the doubling times (DTs) of each mutant increased only modestly with decreasing sequence identity (average increase in DT, 4.6 s per mutation). The three-dimensional structure of the 30S ribosome showed that at least 40.7% (628/1,542) of the nucleotides were variable, even at ribosomal protein-binding sites, provided that the secondary structures were properly conserved. Our results clearly demonstrate that 16S rRNA functionality largely depends on the secondary structure but not on the sequence itself.

functional plasticity

The ribosome plays a crucial role in translating mRNA-encoded information into proteins, which consists of 2 (large and small) subunits. In prokaryotes, the small 30S subunit consists of 16S rRNA and 21 proteins, whereas the large 50S subunit consists of 23S rRNA, 5S rRNA, and 36 proteins. Both in bacteria and archaea, 16S or 23S rRNA shapes the structural core of the subunit particle, whereas many ribosomal proteins are often found on the surface of the subunit, with extensions that protrude into the RNA core (1–3). In contrast, 5S rRNA (small RNA with 120 bases) simply attaches to the 50S subunit and constitutes a structure called a central protuberance (1, 3). Given this structural and mechanistic complexity, as well as their biological significance, it is generally believed that the rRNA genes coevolved with their cognate sets of ribosomal protein genes. In other words, horizontal gene transfer should not occur in rRNA genes because such radical changes would compromise the structural integrity of the ribosome, leading to immediate cell death (4, 5). Taking advantage of this species-specific characteristic (as well as their omnipresence and adequate mutation rates and the adequate amount of information available), researchers have used 16S rRNA genes for the last 30 y to infer the relationships among prokaryotes (4, 6).

Despite the apparent conservative nature of rRNA, some genetic studies using *Escherichia coli* have reported the mutational robustness of rRNA even at protein-binding sites (7, 8). In addition, there is growing evidence to suggest that horizontal gene transfer of 16S rRNA genes takes place in nature (9–13). Recently, using experimental horizontal gene transfer (i.e., the interspecies exchange of *E. coli* 16S rRNA genes with a foreign counterpart), we have also shown that the active hybrid ribosome could be reconstituted in vivo by using *E. coli* Δ7, a null mutant of the ribosomal RNA (*rrn*) operon, as a host (14). Surprisingly, the 16S rRNA from not only the Gammaproteobacteria species *Serratia ficaria* (94.6% identity) but also from the evolutionarily distant Betaproteobacteria species *Ralstonia pickettii* (81.6% identity) have been found to form ribosomes active enough to support the growth of *E. coli* Δ7 (14). All these studies strongly suggest that the ribosome may be much more plastic than had previously been believed, perhaps enough to accommodate foreign 16S rRNA from diverse organisms.

To investigate the mutational robustness of 16S rRNA in this study, we screened for functional genes in *E. coli*. This study uses a functional metagenomic approach (15) to extensively study the possible sequences of RNA-encoding genes. Various 16S rRNA genes were obtained from a diverse lineage of bacteria whose molecular basis for functionality has been discussed based on the primary and secondary structures of 16S rRNA as well as the three-dimensional structure of the ribosome.

Results and Discussion

Sequence Diversity of Functional 16S rRNA Genes in *E. coli*. To rigorously determine the extent to which 16S rRNA genes are interchangeable with those in *E. coli*, we used environmental DNA (metagenome) as a source of the 16S rRNA genes. This approach takes advantage of the large sequence (and the secondary structure's) diversity of the 16S rRNA genes in nature (16) and selects those that are functional in *E. coli*. Genetic selection was carried out as described previously (14) (Fig. S1). Briefly, *E. coli* strain KT101 lacking all seven rRNA operons (*rrnA, B, C, D, E, G, H*) in its chromosome (14, 17) was used as the host; its growth was complemented by the *rrnB* operon encoded in the rescue plasmid pRB101 (pSC101 ori, ampicillin resistant). The plasmid contains the counter selectable marker *sacB* and thus can be conditionally eliminated on addition of sucrose into the medium (14, 17). Using general universal primers (18), we amplified approximately the entire length of the 16S rRNA gene (8–1,541, *E. coli* numbering) from environmental DNA. The amplicons were cloned into another *rrnB*-encoding plasmid pRB103 (pSC101 ori, Zeocin

Author contributions: K.K. and K.M. designed research; K.K., Y.Y., and K.M. performed research; K.K. and Y.Y. analyzed data; and K.K., Y.Y., and K.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the DNA Data Bank of Japan (DDBJ) database, http://www.ddbj.nig.ac.jp (accession nos. AB597523–AB597555).

[1]To whom correspondence should be addressed. E-mail: miyazaki-kentaro@aist.go.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1213609109/-/DCSupplemental.

resistant) (Fig. S1), and an expression library of diverse 16S rRNA genes was created. We then screened this library (~15,000 clones) for functional sequences that complemented the growth of KT101 on sucrose-containing plates.

After counterselection, ~200 clones of KT103-derivatives (carrying pRB103 whose 16S rRNA gene was substituted with foreign genes) were obtained, from which 33 nonredundant 16S rRNA genes (A01–H03) were identified. Through multiple alignment of *E. coli* 16S rRNA and our metagenomically retrieved 16S rRNA sequences, it was found that at least 628 (40.7%) of the 1,542 nucleotides were variable, indicating marked mutational robustness of the 16S rRNA. Strikingly, the functional 16S rRNA sequences (except A10 and F02, which were 99.0% identical to *E. coli* 16S rRNA) obtained in this study showed only 80.9–89.3% identity to *E. coli* 16S rRNA, which was well below the value reported thus far (*Proteus vulgaris* 16S rRNA, 94% identity to *E. coli* 16S rRNA) (19).

The phylogenetic relationships of these functional 16S rRNA genes, as well as those of several relevant bacteria, are shown in Fig. 1*A*. Among the 33 clones, 31 were derived from the diverse lineages of Gammaproteobacteria and two clones (B07 and A09) were from Betaproteobacteria. No clones were obtained from other Proteobacteria classes or other non-Proteobacteria phyla, suggesting the presence of a distinct threshold for 16S rRNA functionality.

The growth phenotypes, defined by the doubling times (DTs) of the mutants, showed a weak negative correlation with sequence identity to the *E. coli* 16S rRNA gene (Fig. 1*B*). The DTs of each mutant increased only modestly with decreasing sequence identity; the average increase in DT was 4.6 s per point mutation. The Betaproteobacteria clone B07, whose 16S rRNA sequence showed the most distant relation to *E. coli* (differing by 274 nucleotides, 80.9% identity), had a DT of 42.0 min, which was still comparable to that of *E. coli* (34.5 min).

Structural Basis for the Mutational Robustness in 16S rRNA. We then questioned why the ribosome was so tolerant to the interspecies exchange of 16S rRNA. From in vitro studies, it has already been observed that 30S and 50S subunits can be exchanged between distantly related species (e.g., *E. coli* and *Bacillus subtilis*), suggesting the sufficient compatibility of bridges between subunits (20, 21). Recently, it has also been reported that the *B. subtilis* 70S ribosome is functional in an *E. coli* in vitro translation system, suggesting the flexibility of interactions between the ribosome and its interacting partners (e.g., initiation factors, elon-

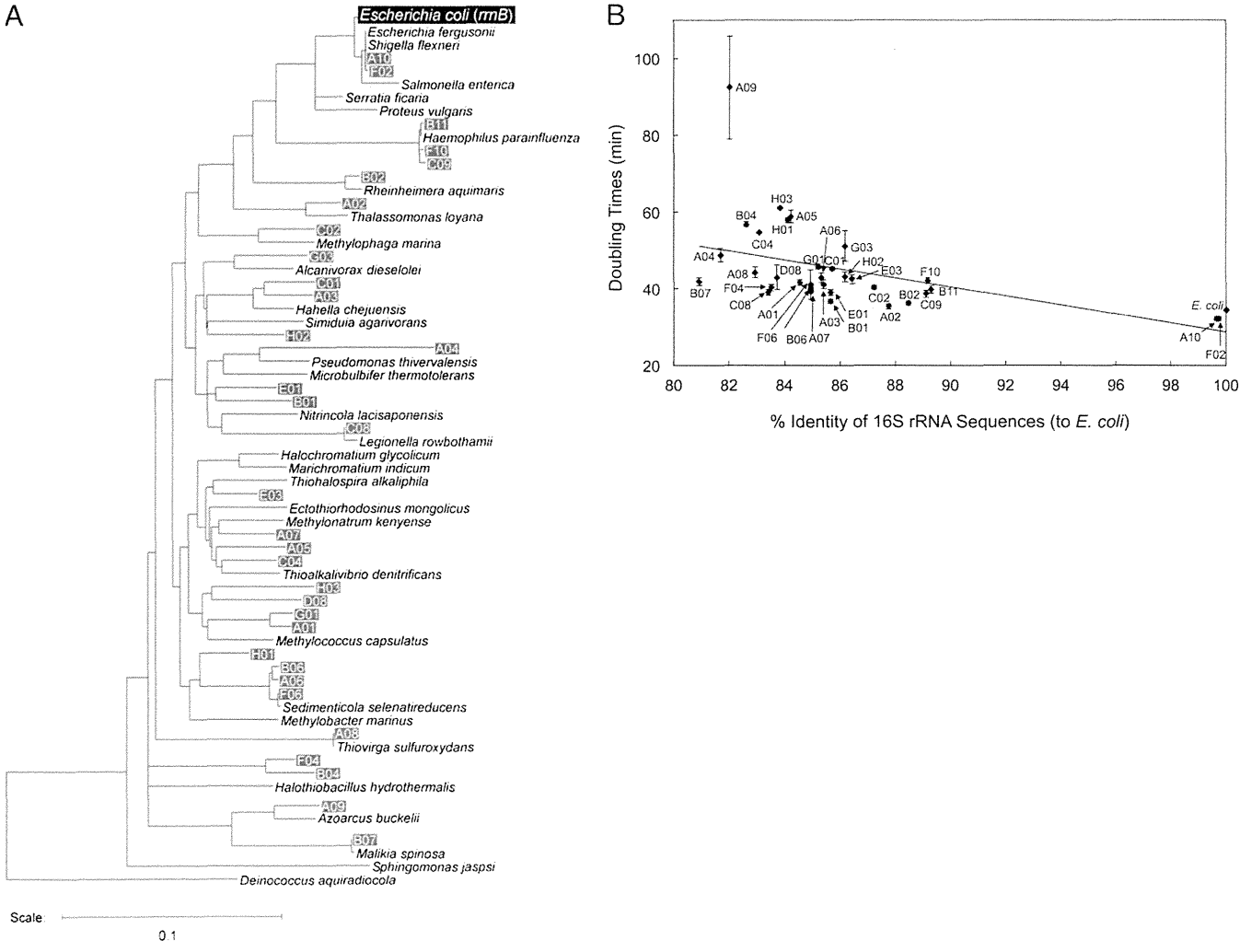

Fig. 1. (*A*) Neighbor-joining phylogenetic tree of the 16S rRNA genes. Environmental 16S rRNA genes that were functional in *E. coli* KT103 are shown as clone ID (A01–H03) with their closest relatives (as nomenclature). Several other relevant strains are also shown. (*B*) DTs (in minutes) versus sequence identity (%, compared with the *E. coli* sequence) of KT103 derivatives harboring foreign 16S rRNA genes. The black slope was fitted to the data. The average increase in DT was calculated to be 4.6 s per point mutation.

gation factors, release factors, and tRNAs) (22). In this article, we have therefore mainly focused on analyzing the interactions between 16S rRNA and ribosomal proteins. The 16S rRNA is typically recognized by ribosomal proteins via salt bridges between phosphate oxygen atoms of the RNA backbone and positively charged protein residues, where nucleotide bases are not strictly discriminated (2, 8). Coincidently, all functional 16S rRNAs, irrespective of their identities (80.9–99.7%), retained their secondary structures; their consensus sequence could easily be superimposed onto the secondary structure map of *E. coli* 16S rRNA (16) (Fig. 2). For example, although the individual sequence of helix (h) 21, which provides a binding site for ribosomal protein S8 (7), varied among functional sequences (0–25.0%) (Fig. S2), compensatory interactions between nucleotide pairs (including both Watson–Crick and noncanonical base pairs) were conserved to form an identical secondary structure (Fig. S3).

To analyze the contribution of the 628 variable nucleotides to ribosomal structure formation in more detail, we constructed contact maps based on the crystal structure of the *E. coli* ribosome (PDB code: 3R8O) (23) (Fig. 3). In these maps, all atomic interactions between oxygen, phosphorus, or nitrogen atoms within 3.4 Å were extracted and dotted in black. These interactions included both specific electrostatic interactions and hydrogen bonds. We then overlaid the points for the 628 variable nucleotides (red dots) onto the RNA–RNA (intra-16S rRNA) (Fig. 3*A*) and RNA–protein (16S rRNA-ribosomal proteins) (Fig. 3 *B* and *C*) contact maps. In the RNA–RNA contact map, the interactions of the compensatory nucleotides within the RNA helices were seen as patterns that protruded perpendicular to the diagonal line (Fig. 3*A*). In this map, 376 nucleotides, representing 79.7% of the total of 472 variable nucleotides, occurred in this protruding pattern, confirming the compensatory and conservative nature of the RNA secondary structure.

In contrast, 133 of the total 410 RNA–protein interactions involved variable nucleotides (Fig. 3*B*). Thus, at least 32.4% of all RNA nucleotides at protein-binding sites are not functionally recognized by ribosomal proteins in a base-specific manner. Specifically, more than half of all interactions involving ribosomal proteins S6 (83.3%), S8 (75.0%), S10 (75.0%), S17 (63.2%), and S2 (50%) did not require base specificity (Fig. 3 *B* and *C*). Moreover, ribosomal protein S20, which connects the 5′ and 3′ minor domains of 16S rRNA, was also base-nonspecific at the 3′ minor domain (Fig. 3*B*). Many residues involved in the binding sites of the ribosomal proteins such as S4 (33.3%), S5 (15.8%), and S12 (8.1%), which are located in the functional center of the 30S subunit (1, 2, 24), were also found to be more or less flexible (Fig. 3 *B* and *C*). These data strongly suggest that the correct formation of the secondary structure of 16S rRNA is critical for the establishment of the 30S subunit's tertiary structure and function; the primary structure of the 16S rRNA itself is less important. This finding is consistent with previous genetic studies (7, 8, 19) but conflicts with the complexity hypothesis (5), which stresses the importance of conserving the primary structure of rRNA.

Overall, we found that the conservation of secondary structures is critical for the functionality of 16S rRNA but with some exceptions; radical changes were found, including some deletions and insertions. Compared with the RNA helices of *E. coli*, h6, h10, and h17 of H01 were shorter by 7, 15, and 25 bases, respectively, whereas h6 of B01 was longer by 13 bases (Fig. S4*A*). However, these changes were found at the tips of the RNA helices that were not involved in protein binding but instead were exposed to the surface, providing basis for the exceptional flexibility (Fig. S4*B*).

Conclusion and Perspectives. The architecture of the ribosome is one of the most complex and sophisticated among that of biological machineries. The assembly process as well as the assembled form is highly complex. This extreme ribosomal complexity is believed to be the basis for the species specificity in 16S rRNA, although solid experimental evidence is not available in this regard (4, 6). It should be noted that 5S rRNA genes, which have less complex structures, are known to be freely exchangeable in bacteria, at least in vitro (25). Using *E. coli* as a host, Fox's group has also shown that 5S rRNA from any *Vibrio* species, except for *Vibrio gazogenes*, could be horizontally transferred (26). In this study, through an evolutionary experiment in *E. coli*, we have tried to clarify the constraints on horizontal gene transfer of 16S rRNA and observed the unexpected plasticity of 16S rRNA. Our finding suggests that ribosomal proteins recognize their binding sites primarily based on the secondary structure of the 16S rRNA. Thus, the establishment of the secondary structure seems to precede the assembly of the ribosome. This base-nonspecific recognition mechanism likely reflects the origin of molecular recognition between RNAs and proteins, which may also apply to the evolution of 23S rRNA. In fact, our preliminary experiment suggested that at least 645 of a total of 2,904 *E. coli* 23S rRNA bases can undergo mutations. This high mutational robustness should provide the basis for the horizontal transfer of 16S rRNA genes observed in some microorganisms in nature (9–13).

By screening mutations that are deleterious to the functionality of the 16S rRNA, Mankin et al. created a list of conservative nucleotides (27) (Fig. 2). Our approach is complementary to their approach but additionally screens directly for mutable nucleotides, taking advantage of the vast sequence diversity of the metagenome (15). The list of variable nucleotides identified in our functional 16S rRNA sequences is highly valuable for mutational studies of the ribosome as the number, positions, and patterns far exceed those reported thus far. For example, the Ribosomal Mutation Database (http://www.oxfordjournals.org/nar/database/summary/229) contains only 321 variable nucleotides. One direct practical application is the use of our mutant strains to evaluate the probability of the emergence of resistance mutations during the development of novel antimicrobial agents that target the ribosomal 30S subunit.

Materials and Methods

Functional Metagenomic Approach. E. coli *strains.* SQ171 (Δ*rrnG* Δ*rrnA* Δ*rrnD* Δ*rrnE* Δ*rrnH* Δ*rrnB* Δ*rrnC*/pTRNA67, pKK3535), a Δ7 prrn strain, was kindly provided by the Suzuki laboratory (University of Tokyo), with the permission of Drs. Selwyn Quan (Stanford University, Stanford, CA) and Catherine L. Squires (Tufts University, Medford, MA), who originally constructed the strain (28) and plasmids (19). The plasmid pKK3535 (*E. coli rrnB*, ampicillin resistant, pBR322 ori) in the SQ171 strain was replaced by pRB101 (*E. coli rrnB, sacB*, ampicillin resistant, pSC101 ori) to generate strain KT101 (14, 17). KT101 was maintained at 37 °C in 2× YT [1.6% (wt/vol) peptone, 1% (wt/vol) yeast extract, and 0.5% (wt/vol) NaCl] medium supplemented with 50 μg/mL ampicillin (14, 17).

Metagenome samples. Genomic DNA, extracted from various environments (soils, fermented products, and seawater), was used as the source of the 16S rRNA genes (15). These samples were mixed and served as polymerase chain reaction (PCR) templates. The mixed metagenomes were found to comprise genomes from diverse lineages of bacteria on the basis of the 16S rRNA sequence analysis: 5.4% from Firmicutes, 5.4% from Spirochetes, 2.7% from Alphaproteobacteria, 32.4% from Betaproteobacteria, 24.3% from Gammaproteobacteria, 8.1% from Deltaproteobacteria, and 21.6% from Epsilonproteobacteria.

Cloning and selection. To amplify the 16S rRNA genes from the environmental metagenome by PCR, we used the KOD-Plus-Ver.2 DNA polymerase (Toyobo) and the primer pairs Bac8f(19A)–UN1541r(1527U) or Bac8f(19C)–UN1541r (1527C) (14, 18). The primer sequences used are as follows: Bac8f(19A), 5′-AGAGTTTGATCATGGCTCAG-3′; UN1541r(1527U), 5′-AAGGAGGTGATCCAACC-3′; Bac8f(19C), 5′-AGAGTTTGATCCTGGCTCAG-3′; and UN1541r(1527C), 5′-AAGGAGGTGATCCAGCC-3′. The resulting amplicons were cloned into pRB103 (containing the entire *rrnB* operon, pSC101 ori, and Zeocin resistance marker) (14, 17) by using the In-fusion procedure (14), and the products were used directly to transform KT101 (Fig. S1). Transformants were then selected on LB [1% (wt/vol) peptone, 0.5% (wt/vol) yeast extract, and 0.5% (wt/vol) NaCl] agar plates containing 100 μg/mL Zeocin (Invitrogen). At this stage, we identified healthy (2 mm diameter after 16 h), midsized, small, and tiny colonies (the great majority indicating a dominant-negative phenotype) at a ratio of ~2:1:1:150, suggesting the functional diversity of the cloned

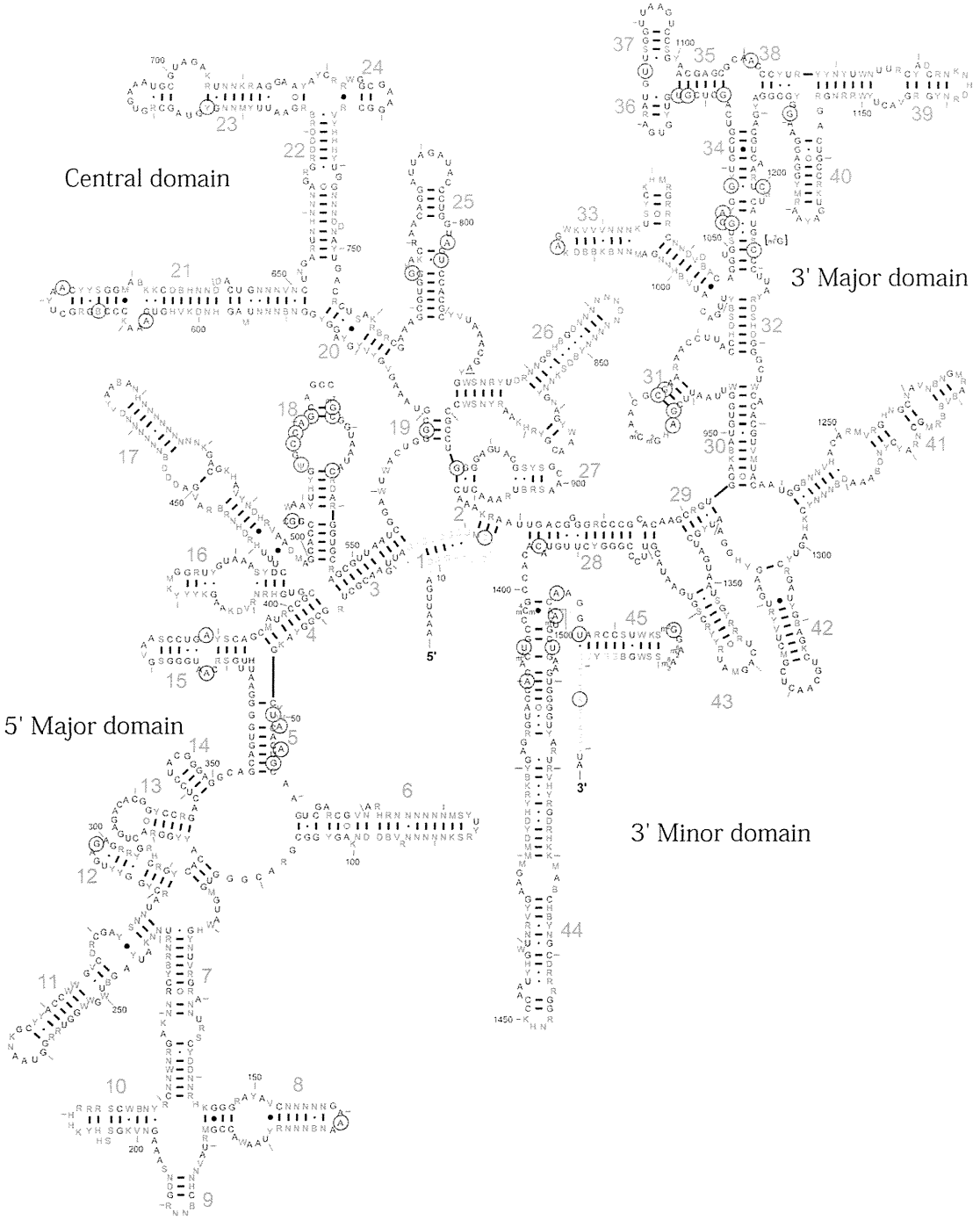

Fig. 2. Variable 16S rRNA nucleotides in *E. coli*. Thirty-three nonredundant functional 16S rRNA sequences that were obtained using a functional metagenomic approach were aligned, and their consensus sequences were superimposed onto the *E. coli* secondary structure map. The original *E. coli* 16S rRNA map was obtained from the Comparative RNA Web site (16). The 628 variable nucleotides are shown in red. Base nomenclature: R can refer to A or G; Y can refer to U or C; W can refer to A or U; K can refer to U or G; B can refer to all bases except A; D can refer to all bases except C; H can refer to all bases except G; V can refer to all bases except U; and N can refer to all bases (A, U, G, or C). The nucleotides in green are the binding sites of the PCR primers—that is, Bac8f at the 5′-end or UN1541r at the 3′-end (14, 18). The nucleotides circled in blue are the invariable nucleotides reported by Mankin's group (27). The number of RNA helices (h1–h45) is indicated in pink.

environmental 16S rRNAs. We then randomly picked ~15,000 colonies, suspended them in LB liquid medium, and spotted them onto LB agar plates containing 100 μg/mL Zeocin and 5% (wt/vol) sucrose (14, 17). After this counterselection, we obtained 89 viable KT103 derivatives, which were colony purified. All of the KT103 derivatives obtained could no longer grow on LB agar plates containing 50 μg/mL ampicillin, indicating complete removal of pRB101.

Phylogenetic Analysis. Plasmid DNA was extracted from the 89 KT103 derivatives, and the entire lengths of the 16S rRNA genes were sequenced. Multiple sequence alignment was carried out using the CLUSTAL W program (http://clustalw.ddbj.nig.ac.jp/top-j.html). A neighbor-joining tree was constructed and drawn using the TreeView software. Because of several near-identical sequences, we excluded all redundant sequences, except one from each group, to yield 33 representative clones. To further characterize the 16S rRNA