Mass spectrometric revival of an L-rhamnose–
and D-galactose–specific lectin from a lost strain of
Streptomyces

Received for publication, August 16, 2017, and in revised form, November 2, 2017 Published, Papers in Press, November 3, 2017, DOI 10.1074/jbc.M117.812719

Yoko Fujita-Yamaguchi†§, Karine Bagramyan‖, Yoshiki Yamaguchi‡, Akemi Ikeda†, Naoshi Dohmae++, Teresa B. Hong*, and Markus Kalkum‡‡

From the Departments of † Molecular & Cellular Biology, ‡ Diabetes Complications & Metabolism, and § Molecular Immunology, Beckman Research Institute of City of Hope, Duarte, California 91010 and the ¶ Structural Glycobiology Team, Systems Glycobiology Research Group, RIKEN Global Research Cluster and the ** Biomolecular Characterization Unit, RIKEN Center for Sustainable Resource Science, RIKEN, Saitama 351-0198, Japan

Edited by Gerald W. Hart

Blood type B-specific Streptomyces sp. 27S5 hemagglutinin (SHA) was discovered and characterized in the 1970s. Although strain 27S5 has been lost, the purified SHA protein survived intact under frozen conditions and retained its activity. Using modern techniques, here we further characterized SHA. Fourier-transform ion cyclotron resonance MS analysis determined the average molecular mass of SHA as 13,314.67 Da. MS of digested SHA peptides, Streptomyces genomic database matching, and N-terminal sequencing solved the 131-residue amino acid sequence of SHA. We found that SHA is homologous to N-terminally truncated hypothetical proteins encoded by the genomes of Streptomyces lavendulae, Streptomyces sp. Mg1, and others. The gene of the closest homologue in S. lavendulae, a putative polysaccharide deacetylase (PDSL), encodes 68 additional N-terminal amino acids, and its C terminus perfectly matched the SHA sequence, except for a single Ala-to-Glu amino acid difference. We expressed recombinant SHA(PDSL-A108E) (rSHA) as an enzymatically cleavable fusion protein in Escherichia coli, and glycan microarray analyses indicated that refolded rSHA exhibits the blood type B– and L-rhamnose–specific characteristics of authentic SHA, confirming that rSHA is essentially identical with SHA produced by Streptomyces sp. 27S5. We noted that SHA comprises three similar domains, representing 70% of the protein, and that these SHA domains partially overlap with annotated clostridial hydrophobic with conserved W domains. Furthermore, examination of GFP-tagged SHA revealed binding to microbial surfaces. rSHA may be useful both for studying the role of SHA/clostridial hydrophobic with conserved W domains in carbohydrate binding and for developing novel diagnostics and therapeutics for L-rhamnose–containing microorganisms.

In 1972, culture supernatants of 333 Actinomyces bacteria strains isolated from the greater Tokyo area were screened for hemagglutination activity to identify microbial lectins (1, 2). The culture supernatant of Streptomyces sp. 27S5 exhibited blood type B-specific activity. This was very unique at the time, because previously known plant lectins were either A- or O-blood type–specific. 60 mg of the identified lectin, named Streptomyces hemagglutinin (SHA), was purified to homogeneity from a 15-liter culture broth of Streptomyces sp. 27S5 using gum arabic affinity chromatography, achieving a 13,300-fold enrichment with 64% recovery of the total activity (3). More than 200 mg of SHA was ultimately purified and subjected to various analyses. SHA was characterized as a small protein (~11 kDa) with unique characteristics, such as rare blood type B specificity, an atypical tryptophan-rich nature, and two carbohydrate-binding sites (3, 4). In the present study, the availability of modern mass spectrometry techniques prompted our further investigation of the archived SHA, which had been kept frozen for 40 years. Unfortunately, the original Streptomyces sp. 27S5 strain was lost, and so was its original genetic makeup. However, a vast database of Streptomyces genomic data has since become available, making it possible to relate the archived SHA lectin to protein amino acid sequences predicted from the genomes of other Streptomyces strains (5).

In the present study, we combined contemporary technologies with information available in the Streptomyces genomic databases to solve the primary structure of SHA, which is the only surviving remnant of the lost Streptomyces strain. Here we

This work was supported in part by funds from the Hermann Foundation (to M. K.). This work was also supported in part by Grants-in-Aid for and Scientific Research B 16H04758 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (to Y. Y.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains Tables S1–S3 and Figs. S1–S3.

1 To whom correspondence may be addressed: Dept. of Diabetes Complications & Metabolism, Beckman Research Institute of the City of Hope, 1500 East Duarte Rd., Duarte, CA 91010. Tel.: 626-218-4669; Fax: 626-301-8280; E-mail: yyamaguchi@coh.org

2 To whom correspondence may be addressed: Dept. of Molecular Immunology, Beckman Research Institute of the City of Hope, 1500 East Duarte Rd., Duarte, CA 91010. Tel.: 626-301-8301; Fax: 626-301-8186; E-mail: mkalkum@coh.org

3 The abbreviations used are: SHA, Streptomyces hemagglutinin; rSHA, recombinant SHA; PDSL, polysaccharide deacetylase; ChW, clostridial hydrophobic with conserved W; ESI, electrospray ionization; FTICR, Fourier transform ion cyclotron resonance; TCEP, tris(2-carboxyethyl) phosphine; RBL, L-rhamnose–binding lectin; CRD, carbohydrate-recognition domain; Trx, thioredoxin; NTA, nitrilotriacetic acid.

© 2018 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.
show that SHA is almost identical (>99%) with the C-terminal domain of a putative polysaccharide decacylase (PDSL) annotated in one version of the *Streptomyces lavendulae* genome, in which it differs by a single Ala-to-Glu amino acid change. Using a glycan array chip, we demonstrated that recombinant SHA(PDSL-A108E) has the same carbohydrate-binding specificity and affinity as the archived SHA purified from *Streptomyces* sp. 27S5, which validated that it is the recombinant SHA (rSHA).

Furthermore, we prepared a green fluorescent protein–SHA fusion protein and demonstrate its binding to microbial surfaces. rSHA can be expressed in different forms for targeting or detecting rhamnose- or galactose-rich glycans on the surface of microorganisms. We expect that rSHA will provide an effective starting material for developing novel diagnostics and therapeutics for galactomannan and L-rhamnose–containing microorganisms.

**Results**

**Preservation of active SHA for more than 40 years**

We first confirmed that the SHA protein that was purified 40 years ago and archived in a frozen state, was intact, readily bound to a gum arabic affinity chromatography column, and specifically eluted with a competing monosaccharide d-galactose or L-rhamnose (Fig. S1, lanes 1–3).

**Determination of the molecular mass of SHA using mass spectrometry**

The molecular mass of SHA was previously estimated to be ~11 kDa, based on various approaches, including gel filtration in the presence of 6 M guanidine hydrochloride, SDS-PAGE, and sedimentation equilibrium analysis (3). To determine the molecular mass of SHA more precisely, we applied electrospray ionization (ESI) Fourier transform ion cyclotron resonance (FTICR) MS. This high-resolution mass spectrometric technique revealed a precise average molecular mass of 13,314.67 Da, a monoisotopic mass of 13,306.65 Da, and the presence of a covalently attached hexose in ~25% of the SHA molecules (Fig. 1).

**Identification of SHA homologues in Streptomyces genomes**

To determine the sequence identity of SHA, we conducted bottom-up proteomics experiments. We digested SHA separately with several proteases to generate overlapping peptides. These peptides were then analyzed by LC coupled with high-
L-Rhamnose–specific lectin with three repeated domains

resolution multistage mass spectrometry (MS/MS). An initial database search was performed in April 2014 and revealed a closely matching SHA homologue in the genome of *Streptomyces* sp. Mg1 as a hypothetical protein (GenBank™ accession no. EDX26679.1; Ref. 6; data not shown). By July 2014, more *Streptomyces* genome sequences became available; subsequently, a more refined search led to the identification of a homologue in *S. lavendulae* with even better scores for MS/MS database matching. The digested SHA peptides aligned almost completely with the deduced C-terminal amino acid sequence of the putative PDSL of *S. lavendulae* (accession no. WP_051840348.1; Fig. S2). SHA matched with the C-terminal 131 amino acids of the hypothetical 199-amino acid PDSL protein, except for a partial sequence stretch consisting of nine amino acids from SHA positions 101–109. However, the mass spectrometric data did not cover any sequence of the N-terminal portion of either the *Streptomyces* sp. Mg1 or *S. lavendulae* protein, comprising 74 and 68 amino acid residues, respectively.

**Determination of the N-terminal amino acid sequence of SHA**

Previous amino acid sequencing of reduced and carbamylated SHA revealed the N-terminal amino acids to be AT-VCYAA (7), where X indicates an undetermined residue. To confirm these results and identify additional amino acids in the sequence, we performed N-terminal sequencing of the archived SHA. We identified ~30 amino acids to be ARTVcYAAHVE-GIGWQGAVcDGAVA (lowercase letters indicate tentative identification). Together, the two independent N-terminal sequencing results strongly suggested that the N-terminal sequence of SHA was ARTVCY.

**Solution of the primary structure of SHA**

By considering the N-terminal sequencing information, the molecular mass of the intact SHA protein, and the database matching with digested peptides, we concluded that the SHA sequence had to be almost identical to the C-terminal portion of the PDSL, residues 69–199. To identify how the amino acid sequences differ between SHA and the SHA domain of PDSL, SHA(PDSL), we generated a recombinant thioredoxin (Trx)-SHA(PDSL) fusion protein and compared peptide fingerprints of SHA and SHA(PDSL). First, we cloned the homologous SHA domain from PDSL into a PET32 vector to transform *Escherichia coli*, from which the fusion protein was purified using nickel-NTA resin (Fig. S1, lane 4). Then we digested the purified recombinant Trx-SHA(PDSL) fusion protein using multiple enzymes to generate overlapping peptides for LC-MS and MALDI-MS analyses, as for SHA above. Finally, we compared the LC-MS/MS data sets of the digested peptides from SHA and its homologue SHA(PDSL). We found that the sequence of SHA differed from that of SHA(PDSL) by a single A108E change (Fig. 2, A, B, and D). This was confirmed by calculating and comparing the molecular masses for SHA and SHA(PDSL) as a 58-Da mass difference. Thus, we concluded that the amino acid sequence of archived SHA is that of SHA(PDSL-A108E), which represents the SHA domain from PDSL with a single A108E change. For validation, we are providing a complete list of all LC MS/MS peptide spectrum matches for archived SHA after individual digestions with ArgC, trypsin, and chymotrypsin in Table S1.

To determine disulfide bonds in SHA, we used endoprotease ArgC to digest archived SHA, with or without reduction with tris(2-carboxyethyl) phosphine (TCEP), followed by high-resolution Orbitrap LC-MS. Comparison of the spectra of two digested peptides before and after TCEP reduction showed a clear 2-Da mass difference (Fig. 2C). This indicates that SHA contains two consecutive disulfide bonds that connect cysteine residues Cys5 with Cys20 and Cys48 with Cys63, as illustrated in Fig. 2E. No other disulfide bond–connected peptides were detected. Taken together, these results allowed us to deduce the primary structure of SHA (illustrated in Fig. 2E and summarized in Tables 1 and 2, along with repetitive domain structures).

**Characterization of the carbohydrate-binding properties of SHA**

In contrast to archived SHA, the first Trx-SHA(PDSL) fusion protein we generated (above) was poorly soluble and not suitable for functional analyses. Therefore, we expressed a novel construct that encoded an *E. coli* codon-optimized, His-tagged, Trx-SMT3 (SUMO family protein)-SHA(PDSL-A108E) fusion protein. This protein was refolded on the nickel NTA column in the presence of D-galactose, and soluble rSHA was cleaved off from the Trx-SMT3 portion using His-tagged Ulp1 (8). The purified rSHA is shown in Fig. 3 (inset) together with archived SHA. To determine the carbohydrate-binding specificity of SHA and rSHA, we used the commercially available glycan array 100 slides, on which 100 synthetic glycans of the most frequently identified structures in the literature are mounted in each of four subarrays. We measured glycan binding using two distinct concentrations of biotin-labeled SHAs (1 ×: 20 µg/ml and 0.1 ×: 2 µg/ml) in the absence and presence of 0.2 M l-rhamnose as a competitive inhibitor. Both SHA and rSHA showed identical carbohydrate-binding specificities (Fig. 3, A and B, and Fig. S3). We observed strong signals for SHA/rSHA binding to β-Gal; α-Rha; Gal-α-1,3-Gal-β-1,4-GlcNAc-β; Gal-α-1,3-Gal-β-1,4-Glc-β; Gal-α-1,4-Gal-β-1,3-GlcNAc-β; Gal-α-1,3-(Fuc-α-1,2)-Gal-β (blood B antigen trisaccharide); Gal-α-1,3-(Fuc-α-1,2)-Gal-β,1,4-Glc-β (blood B antigen tetrasaccharide); and Gal-α-1,3-Gal-β. We observed weaker, but still significant, binding to Gal-α-1,4-Gal-β,1,4-Glc-β; GalNAc-β-1,3-Gal-β,1,4-Glc-β; and Gal-α-1,4-Gal-β,1,4-GlcNAc-β. SHA/rSHA did not bind to the remaining 89 chip-immobilized glycans (Table S2).

In the presence of l-rhamnose, SHA/rSHA binding to all glycans, except α-Rha, was competitively inhibited (Fig. 3, A and B). These results are consistent with earlier hemagglutination inhibition or equilibrium dialysis observations in which SHA bound to l-rhamnose with a higher affinity than to D-galactose (2–4).

**Sequence comparison of SHA and putative homologues**

We identified SHA homologues not only in the *Streptomyces* genome, but also in the genomes of other microorganisms. In total, 11 putative SHA homologues with more than 50% homology to the SHA sequence were identified as N-terminally trun-
cated hypothetical proteins in the genomes of *S. lavendulae*, *Streptomyces* sp. Mg1, *Streptomyces* xanthophaeus, *Streptomyces* sp. Wm6378, *Streptomyces clavuligerus*, *Streptomyces scabiei*, *Streptacidiphilus melanogenes*, *Lentzea* sp. DHS C013, Actinobacteria bacterium, and *Nocardiopsis* sp. NRRL S-836 (Fig. 4A). The N-terminal sequence of the putative SHA homologues varied among homologues, and a corresponding sequence was absent in SHA. In contrast, the C-terminal domain was conserved between SHA and its homologues. Compared with the 131 amino acids of the SHA sequence, the SHA homologues contained 15–133 additional amino acids at the N-terminal end, for a total of 172–265 amino acids.

To compare protein and DNA sequences of SHA and its homologues, we generated a phylogenetic tree (Fig. 4B). Protein sequence homology ranged from 51 to 99%. In the absence of SHA genetic information, we used *S. lavendulae* DNA (438 bases) as the reference query for SHA homologues. DNA sequence homology ranged from 67 to 82%.

The primary sequence of SHA is principally made up of three homologous “SHA domains,” each consisting of 29–33 amino acids (Table 1). Sequence identity between the three SHA domains ranged from 60 to +5 as indicated. D, Orbitrap MS/MS collision-induced dissociation spectrum and fragmentation scheme of a semitryptic SHA peptide providing evidence for Glu108 residue, parent ion m/z 1032.00, z = +2. E, sequence and primary structure of SHA derived from the results in A–D and Table S1. Digested peptides of SHA identified by MS analyses (A, B) are indicated by blue lines; S–S bonds identified by Orbitrap LC-MS (comparisons within C) are indicated by red linking lines; the Ala-to-Glu variation at SHA position 108 (indicated by blue labeled peak in A and by blue linking lines in B) is highlighted in red.
### L-Rhamnose-specific lectin with three repeated domains

#### Table 1

| Location | Amino acid sequence | Number of residues | ChW-like domains |
|----------|---------------------|--------------------|------------------|
| 1–7      | ARTVCCYA            | 7                  |                  |
| 8–36     | AHVDEIQGQVQGACDGA   | 29                 | SHA domain 1     |
| 37–50    | AVTVGQGVGSCMA       | 14                 |                  |
| 51–83    | AHIADIGQGWAADGKAVTV | 33                 | SHA domain 2     |
| 84–96    | LGFQNYSVQRA         | 13                 |                  |
| 97–126   | AVIATSGTGGVCAN      | 30                 | SHA domain 3     |
| 127–131  | VRTDW               | 5                  |                  |

#### Discussion

This study demonstrates that archived SHA produced by *Streptomyces* sp. 27S5 and purified 40 years ago remained intact and maintained its carbohydrate-binding and hemagglutination (data not shown) activities and that the molecular mass and primary structure of the archived SHA were successfully determined using modern mass spectrometric/proteomic strategies. The amino acid sequence of SHA was partially determined by Edman degradation methods in the 1970s, as described in the thesis of Y. F.-Y. (7). That study found redundancy in the N-terminal amino acid sequences of BrCN-cleaved SHA peptides, which was reasoned to be due to the presence of microheterogeneity in the purified protein. The primary structure determined in the current study clearly reveals that the difficulty of sequencing SHA in the 1970s was due to the three homologous SHA domains, which occupy 70% of the SHA molecule. It is fortuitous that the putative PDSL gene of *S. lavendulae* was found in the *Streptomyces* genome database, which was expanded within 2 months after this protein was first revisited after 40 years. Consequently, the primary structure of SHA was revealed at last.

FTICR-MS revealed an average molecular mass of 13,314.67 Da and the presence of a covalently attached hexose in ~25% of the SHA molecules. The MS results suggest that hexose may be a component of SHA. Glycation of Lys in macromolecules, including hemoglobin, serum albumin, crystalline, and collagen, has been well-studied (12–14). Given that the original SHA was obtained from a culture medium containing 2% D-fructose, it is possible that D-fructose was non-enzymatically attached to ε-amino groups of Lys. SHA was exposed to a high concentration of D-galactose after the original affinity purification, and we found significant amounts of D-galactose remaining in the archived SHA sample. Thus, it is also possible that D-galactose present in the SHA solution may have caused such a covalent linkage. Alternatively, it is possible that the hexose was added post-translationally by *Streptomyces*. However, the mass spectrometric data indicated multiple hexose-modified residues and not a single defined site (data not shown), hinting the presence of an inhomogeneous chemical reaction rather than a well-defined in vivo post-translational modification.

We solved the 131-amino acid primary structure of SHA by showing that peptides derived from SHA aligned to the C-terminal two-thirds of the hypothetical protein from *S. lavendulae* (PDSL) with >99% identity. Close comparison of peptides derived from SHA and the SHA domain of PDSL (from now on referred to as SHA(PDSL)) revealed a single amino acid substitution at the SHA-equivalent position 108 in PDSL, from Ala to Glu. Recombinant SHA(PDSL-A108E) showed the same carbohydrate-binding specificity and similar affinity for l-rhamnose as archived SHA. These results confirmed that unmodified archived SHA is identical to SHA(PDSL A108). The SHA (PDSL-A108E) gene was used to express recombinant SHA proteins in different forms, including GFP–SHA. After the confirmation of l-rhamnose and D-galactose glycan specificity of the SHA(PDSL-A108E) protein, we designated this protein as rSHA.

We showed that SHA and 11 hypothetical protein homologues have three ChW-like SHA domains. To date, ChW domains have been exclusively found in the *C. acetobutylicum* species. The three ChW-like domains we identified in SHA and its homologues represent additional examples for non-*C. acetobutylicum* proteins containing ChW domain repeats. The ChW domain is 45–47 amino acids long and features an absolutely conserved tryptophan and high contents of hydrophobic and small amino acids. SHA homologues contain five conserved tryptophan residues, four of which are located in the three ChW-like SHA domains. Like the three SHA domains in SHA, the ChW domains cluster into groups of threes, which
suggests they function as a triplet (10). Although carbohydrate recognition functions have been suggested (9), no conclusive study has been published as to the role of ChW domains. Our rSHA will provide a unique opportunity for studying the role of ChW and SHA domains in carbohydrate binding.

The identified tryptophan residues may be involved in the binding function of SHA. We previously reported that the CD spectrum of SHA strongly resembled that of poly(L-tryptophan) and speculated that tryptophan side chains contributed to a positive CD band at 226 nm (3). We also suggested a potential involvement of tryptophan residues in L-rhamnose binding to SHA (4). Those conclusions were based on solvent-perturbation studies, which demonstrated that the number of solvent-exposed Trp (or average extent of exposure) was two in the absence of L-rhamnose and three in the presence of L-rhamnose. This suggested that one tryptophan residue appears outside as a result of SHA binding to this sugar. Oxidation of two tryptophan residues with N-bromosuccinimide led to complete loss of its carbohydrate-binding activity, which also indicated that these tryptophan residues are important for retaining this activity (4). Using NMR, the current study confirmed the involvement of tryptophans in the binding of SHA to L-rhamnose. In analogy, another L-rhamnose–specific protein, \( \alpha-L-\)rhamnosidase of \( \text{Streptomyces avermitilis} \), has three tryptophan residues binding to L-rhamnose via hydrophobic interaction to the pyranose ring of the sugar (15).

The above-mentioned structural information is a prerequisite for understanding specificity and affinity of SHA. It is essential, however, to carry out extensive binding assays of SHA against a variety of glycans. In this study, we compared the specific binding of both archived SHA and rSHA side by side, using the Glycan Array 100 and comparing two concentrations of SHA in the absence and presence of L-rhamnose. Although semiquantitative, the results clearly revealed the following: 1) SHA bound to D-galactose and glycans containing Gal-\( \alpha-1\rightarrow3 \), which is the key signature of blood type B specificity, as well as

![Figure 3](image-url)

**Table 2** Homology between ChW and SHA domains

| ChW Domain | Homology to C. acetabutylicum ATCC 824 protein Q97E41, found using SMART (simple modular architecture research tool). Key signature of ChW domains, tryptophan (W), is underlined. |
|---------------------------|-------------------------------------------------------------------------------------------------|
| SHA domain 1              | AHVQNIQGDVQYSNAGAEGTGDSPGQVALRELRENNP                                                         |
| SHA domain 2              | AHLLDQGQNAADDAG-KAVTVGQGDSRREAA                                                              |
| SHA domain 3              | AHVA-DYGRLN---ARGNPVYGGTGQGSRREAA                                                             |

![Inset](image-url)
L-rhamnose-specific lectin with three repeated domains

It is interesting to note that L-fucose does not bind to SHA (Table S2). Likewise, carbohydrates with terminal L-fucose that lack terminal D-galactose do not bind SHA (Table S2). L-Rhamnose and L-fucose are both deoxy hexoses. L-Rhamnose is 6-deoxy-L-mannose, whereas L-fucose is 6-deoxy-L-galactose. Comparing the structures of these carbohydrates reveals that the orientation of the hydroxyl groups at positions C2 and C4 differs. This suggests that the C2 and/or C4 hydroxyl groups of L-rhamnose are essential for SHA binding. L-Rhamnose and D-galactose share a common architecture from C2-OH to C4-OH, which may form a critical SHA binding motif. As previously shown, D-galactose is a good inhibitor for SHA binding, whereas 2-deoxy-D-galactose, galactosamine, and GalNAc are not (2). This suggest that C2-OH is important for SHA binding, consistent with our new data.

Gum arabic has been effectively used to purify SHA and SHA fusion proteins in the past and this study. As we previously reported (2), hemagglutination of type B-erythrocytes by SHA was inhibited in the presence of plant-originated galactomannans, with guar gum / locust bean gum / gum arabic. Gum arabic is known to contain galactose, rhamnose, and arabinose as major components (16). However, the precise substructure required to bind SHA remains to be determined.

Although microbial lectins with similar characteristics to SHA have not been reported, significant data on L-rhamnose-binding lectins (RBLs) from fish eggs are available (17–21). Interestingly, RBLs from a number of different fish species are composed of two or three domains consisting of ~100 amino acids, which are known as carbohydrate-recognition domains (RBL CRDs) (22, 23). A lectin purified from sea urchin (Anthocidaris crassispina) eggs (SUEL) was reported to contain a
galactose-binding lectin domain (24) but was later shown to bind to \(\alpha\)-rhamnose preferentially, which seems reasonable given that \(\alpha\)-rhamnose and \(\beta\)-galactose share the same hydroxyl group orientation at C2 and C4 of the pyranose ring structure (22, 23). The RBL CRD, also called SUEL-type lectin domain, is composed of eight highly conserved half-Cys and several other conserved segments, e.g. YGA in the N-terminal and DP and K in the C-terminal domain (22). However, RBL CRD shows no homology to SHA domains, because of its domain size, which is over three times longer than the SHA domain; the absence of tryptophan, which is the signature of SHA domains; and its heavily disulfide-linked domain structure. Although RBL CRD tandem motifs are somewhat architecturally similar to the three SHA domain repeats presented in this study, three-dimensional structures and the amino acid residues required for \(\alpha\)-rhamnose binding must be solved before speculating on the contribution of RBL CRD and SHA domain repeats to carbohydrate binding.

Figure 5. \(^1\text{H}\) NMR analysis for \(\alpha\)-rhamnose binding to archived SHA. \(\alpha\)-Rhamnose was added to the SHA solution from 1 to 5 equivalents. The indole NH region (left panel) and methyl region (right panel) of \(^1\text{H}\) NMR spectra are shown. Dashed red lines are included for alignment. Chemical shift changes observed for specific peaks indicate binding of \(\alpha\)-rhamnose to SHA.

Figure 6. Top left panel, representative fluorescent micrographs of \(L.\) casei (Shirota) cells stained with recombinant GFP–SHA fusion protein. Top right panel, recombinant GFP was used as a negative (neg.) control. DAPI counterstaining is shown in blue in the merged micrographs (bottom).
L-Rhamnose–specific lectin with three repeated domains

The functions of L-rhamnose–specific lectins are of particular interest. One suggested physiological role of fish egg lectins is as a defense mechanism against pathogenic bacteria (17). Rhamnose-binding lectins from salmon and trout are involved in innate immunity and recognition of LPSs or lipoteichoic acid, respectively, on the cell surface of bacteria (20, 25). In contrast to animal lectins, lectins produced by microorganisms have different functions. Bacterial surface agglutinins with mannose specificity play roles in cell–cell interactions, as well as in microbial pathogenicity (26). The related functions of SHA are expected to include interactions with outside cells, such as attaching to neighboring plants and surrounding microorganisms, in addition to potential defense mechanisms. It is interesting to note that the closest SHA homologue was found in the S. lavendulae genome encoding a putative PDSL. If expressed by Streptomyces, this enzyme would be expected to catalyze the N- or O-deacetylation of acetylated sugars on the membranes of Gram-positive bacteria. However, it is not likely that SHA has such deacetylation activities, because SHA does not seem to recognize N-acetylated carbohydrates, as seen in the glycan array results.

Our comparison of SHA to genomically derived hypothetical proteins revealed the intriguing observation that the SHA-homologous domains of all 11 hypothetical proteins are localized in the C-terminal regions of the larger ORFs. Under the cultural conditions described (3), we did not observe expression of the SHA-homologous proteins encoded by the genomes of S. lavendulae and Streptomyces sp. Mg1 (data not shown). In contrast, when the original study was performed in the 1970s, three HA activity-positive strains were identified from the 333 Actinomycetales culture supernatants screened (1). Our recent findings generated the next important question, how SHA was expressed and secreted from the lost Streptomyces sp. 27S5. We hypothesize that SHA could have been expressed as a precursor protein with an unknown N-terminal sequence, a signal sequence, and a protease-processing site, so that SHA molecules could be found in the culture broth, as observed 40 years ago.

Our current efforts include expression of soluble recombinant SHA molecules in quantity for three-dimensional structural analyses and regaining an Streptomyces sp. 27S5 equivalent strain for studying the promoter region of the SHA gene, the biosynthesis and secretion of SHA, and the functions and roles of SHA in nature. As we showed that the recombinant GFP–SHA binds to L. casei Shirota cells, a variety of other bacteria and fungi could easily be screened in the future to identify microorganisms that may interact with SHA. A similar approach was reported for a recombinant horseshoe crab plasma lectin that recognizes specific pathogen-associated molecular patterns of bacteria through L-rhamnose (27). In conclusion, despite the lack of genetic information, we have revived a valuable lectin from a lost strain of Streptomyces using protein engineering based on mass spectrometric data.

Experimental procedures

Materials

S. lavendulae strain NCIB 6959/ATCC 14158 and HEK293S cells were purchased from ATCC (Manassas, VA). Streptomyces sp. strain Mg1 was a kind gift from Dr. Paul Straight of Texas A&M University (6). E. coli C41(DE3) and E. clone® were from Lucigen (Middleton, WI). Gum arabic was purchased from Sigma–Aldrich. MS-grade Trypsin, LysC, ArgC, V8 protease, and pepsin were from Promega (Madison, WI). Chymotrypsin was from Worthington Biochemical (LakeWOOD, NJ). pET32b and pcDNA3.1 vectors were from Merck Millipore and Thermo Fisher Scientific, respectively.

Purification and characterization

SHA was purified 40 years ago as described (3) and kept frozen at –80 °C. The purity and quality of the archived SHA were determined using SDS-PAGE. The N-terminal amino acid sequence of SHA was determined using Edman degradation performed on the Procise 494HT Protein Sequencing System (Applied Biosystems, Thermo Fisher Scientific).

Specific binding of SHA to gum arabic gels

Gum arabic gels were prepared according to published methods (3). The archived SHA as well as recombinant SHA proteins were applied to the gum arabic gel column. After washing the column, SHAs were eluted with either 1 M D-galactose in the presence of 1 M NaCl as described (3) or 0.2 M l-rhamnose in the presence of 1 M NaCl.

NMR titration study

NMR analysis was performed using a DRX-500 spectrometer equipped with a cryogenic TXI probe (Bruker BioSpin, Billerica, MA). The probe temperature was set to 298 K. Archived SHA (0.1 mg) was dissolved in 500 μl of 20 mM sodium phosphate buffer, pH 6.5 (H2O:D2O = 9:1). L-Rhamnose solution was added to the SHA solution at molar ratios from 1:0 to 1:5 (SHA:L-rhamnose). Data processing and analysis were performed using XWIN-NMR (version 3.5, Bruker BioSpin). NMR spectra were displayed with XWIN-PLot (version 3.5, Bruker BioSpin).

Mass spectrometry

To determine molecular mass, the intact archived SHA was analyzed using ESI FTICR MS on a Thermo LTQ FTICR (Thermo Fisher Scientific) at ~500,000 resolution. To determine the amino acid sequence of SHA, overlapping SHA peptides were obtained by performing separate enzymatic digestions with trypsin, chymotrypsin, LysC, ArgC, V8 protease, and pepsin and were analyzed by LC-MS on an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher Scientific), as well as by MALDI-MS on a SimulTof Combo 200 instrument (SimulTOF Systems; Virgin Instruments, Marlborough, MA). MS and MS/MS collision-induced dissociation fragmentation data from these peptides were analyzed with Xcalibur software (Thermo Fisher Scientific) and with PEAKS Studio software (Bioinformatics Solutions Inc., Waterloo, Canada). SHA disulfide bond determination was made using MALDI-MS and high (120,000) resolution Thermo Orbitrap Fusion Tribrid mass spectrometer analysis of the intact protein and the digested protein, before and after reduction with 50 μM TCEP, pH 2.0, at 80 °C for 30 min.
Expression of an SHA homologous recombinant protein

We expressed the SHA homologous domain of PDSL from *S. lavendulae*, which showed the highest homology to SHA (>99% identity), as a recombinant protein. To develop this recombinant SHA homologue, a synthetic gene expressing a wild-type SHA of PDSL and a mutant SHA gene with an Alanine-to-Glu amino acid substitution at position 108 (A108E) were produced using *E. coli* codon-optimized overlapping oligo DNA primers and cloned into pET32b (Table S3). The recombinant wild-type SHA(PDSL) was expressed in *E. coli* C41(DE3) as a Trx fusion protein with His-tag. Trx-SHA(PDSL) was purified from *E. coli* cell pellets derived from a 2-liter culture by solubilization and affinity purification on a nickel-NTA resin (Thermo Fisher Scientific). The purified wild-type SHA(PDSL) was digested with multiple enzyme combinations, as described above for SHA, to compare resulting peptides from both proteins.

Because of solubility issues, various fusion proteins of the recombinant SHA(PDSL-A108E) were prepared and expressed in *E. coli*. Of those, a yeast SUMO(SMT3) fusion protein was successfully purified for comparing carbohydrate-binding specificity with that of archived SHA. Briefly, SMT3-fused SHA(PDSL-A108E) was prepared by insertion at the SMT3 and Ulp1 cleavage sites of pET32b/SMT3. *E. coli* (Lucigen) was transformed by pET32b/SMT3-SHA(PDSL-A108E). SMT3-SHA(PDSL-A108E) was purified using a His6 tag–specific nickel-NTA column from transformed cells after solubilization with 5 M urea/B-Per lysis buffer (Pierce), followed by refolding in the presence of 1 M galactose and 10 mM β-mercaptoethanol and eluting with 400 mM imidazole. GFP–SHA was concentrated using Centricon YM10 centrifugal filters (Fisher Scientific) and purified by FPLC with Superdex 75G (GE Healthcare).

*L. casei* Shirota cells were isolated from commercially available Yakult yogurt drink. The authenticity of *L. casei* Shirota was verified by Sanger sequencing of its 16S rRNA by showing 100% match to the reference sequence AB531131. 400 ml of Difco™ *Lactobacilli* MRS broth (Fisher Scientific) was inoculated with 0.1 mg/ml of rSHA (0.1 g/ml) and harvested at 37 °C, harvested by centrifugation, and washed three times with PBS. The cells were resuspended in 5 ml of 70% ethanol and incubated at 22 °C for 30 min under continuous rotation. The cells were washed three times with PBS and finally resuspended in 1 ml of PBS containing 10% glycerol. Bacterial cells were counterstained with DAPI (3 μM) and examined using a Zeiss Observer II system (Carl Zeiss). Fluorescent images were analyzed using Image-Pro Plus and ZEISS ZEN software (Carl Zeiss).

Glycan microarray analyses

Microarray analysis was performed according to the manufacturer’s recommendations using RayBio Glycan Array 100 (RayBiotech, Norcross, GA) slides. Each slide contains four submicroarrays printed with 100 synthetic glycans. Briefly, 200 μl of 0.1 mg/ml of both archived SHA and rSHA were dialyzed overnight at 4 °C against 1× PBS dialysis buffer to avoid contaminating samples with amines prior to biotinylation. Dialyzed samples were incubated with biotin-containing reaction solution at 22 °C for 30 min. Subarrays were blocked for 30 min at 22 °C. After biotinylated SHA samples were diluted with 1× PBS, 400 μl of each sample was added to each subarray. Slide 1 subarrays were incubated with 400 μl of 20 μg/ml (1×) or 2 μg/ml (0.1×) SHA, in the absence or presence of 0.2 M L-rhamnose. Slide 2 subarrays were incubated with 400 μl of 20 μg/ml (1×) or 2 μg/ml of rSHA (0.1×) in the absence or presence of 0.2 M L-rhamnose. The slides were incubated for 16 h at 4 °C for highest intensities. Washing was performed according to the manufacturer’s protocol, followed by incubation with Cy3 dye-conjugated streptavidin. The slides were incubated at 22 °C for 1 h with gentle shaking and then washed multiple times as recommended. The signals were visualized using an Agilent DNA microarray scanner (model G2505C; Agilent, Santa Clara, CA) at 532 nm for Cy3. As controls, separate microarray chips were incubated with either biotinylated concanavalin A or GFP. Specific signal patterns were obtained for concanavalin A, and no signals (other than for the positive controls) were obtained for biotinylated GFP (data not shown). Data extraction and analysis were performed after subtraction of the background and normalization to the internal references provided by the manufacturer, using an ImageJ protein array analyzer software (28).

Staining of *L. casei* (Shirota) cells by fluorescently labeled SHA

Recombinant GFP–SHA was expressed by inserting the SHA(PDSL-A108E) gene at the C terminus of GFP in pET28/GFP, followed by transformation of *E. coli* cells. GFP–SHA was purified from cell pellets collected from 4 liters of culture, after solubilization with 5 M urea/B-Per lysis buffer (Pierce), using a His6 tag–specific nickel-NTA column, followed by refolding in the presence of 1 M galactose and 10 mM β-mercaptoethanol and eluting with 400 mM imidazole. GFP–SHA was concentrated using Centricon YM10 centrifugal filters (Fisher Scientific) and purified by FPLC with Superdex 75G (GE Healthcare).

*L. casei* Shirota cells were stained with fluorescently labeled SHA. L. casei Shirota was stained with 0.1 mg/ml of SHA and 0.1 mg/ml of rSHA were dialyzed overnight against 1× PBS dialysis buffer to avoid contaminating samples with amines prior to biotinylation. Dialyzed samples were incubated with biotin-containing reaction solution at 22 °C for 30 min. Subarrays were blocked for 30 min at 22 °C. After biotinylated SHA samples were diluted with 1× PBS, 400 μl of each sample was added to each subarray. Slide 1 subarrays were incubated with 400 μl of 20 μg/ml (1×) or 2 μg/ml (0.1×) SHA, in the absence or presence of 0.2 M L-rhamnose. Slide 2 subarrays were incubated with 400 μl of 20 μg/ml (1×) or 2 μg/ml of rSHA (0.1×) in the absence or presence of 0.2 M L-rhamnose. The slides were incubated for 16 h at 4 °C for highest intensities. Washing was performed according to the manufacturer’s protocol, followed by incubation with Cy3 dye-conjugated streptavidin. The slides were incubated at 22 °C for 1 h with gentle shaking and then washed multiple times as recommended. The signals were visualized using an Agilent DNA microarray scanner (model G2505C; Agilent, Santa Clara, CA) at 532 nm for Cy3. As controls, separate microarray chips were incubated with either biotinylated concanavalin A or GFP. Specific signal patterns were obtained for concanavalin A, and no signals (other than for the positive controls) were obtained for biotinylated GFP (data not shown). Data extraction and analysis were performed after subtraction of the background and normalization to the internal references provided by the manufacturer, using an ImageJ protein array analyzer software (28).

Author contributions—Y. F.-Y. and M. K. designed the study and wrote the paper. K. B., Y. Y., A. L., and T. B. H. analyzed SHA using NMR, MS, and other biochemical methods. N. D. determined the N-terminal amino acid sequence of SHA. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments—We thank Dr. Paul Straight of Texas A&M University for the kind gift of Streptomyces sp. Mg1. We also thank Dr. Alexander Pozhitkov for helping with microarray reader setup and Dr. Sarah T. Wilkinson for editing the manuscript. Y. F.-Y. thanks the late Drs. Kunio Oishi, Koichi Suzuki, and Kazutomo Imahori for the excellent mentoring and valuable comments she received while conducting the initial phase of this research at the University of Tokyo Graduate School. City of Hope facilities, including the Integrative Genomics Core used in this study, were supported, in part, by the NCI, National Institutes of Health under Grant P30 CA33572.
**L-Rhamnose–specific lectin with three repeated domains**

**References**

1. Fujita, Y., Oishi, K., and Aida, K. (1972) Hemagglutination by culture broth of Actinomycetes and Aspergillus. *J. Gen. Appl. Microbiol.* **18**, 73–75 CrossRef

2. Fujita, Y., Oishi, K., and Aida, K. (1973) Sugar specificity of anti-B hemagglutinin produced by Streptomyces sp. *Biochem. Biophys. Res. Commun.* **53**, 495–501 CrossRef Medline

3. Fujita, Y., Oishi, K., Suzuki, K., and Imahori, K. (1975) Purification and properties of an anti-B hemagglutinin produced by Streptomyces sp. *Biochimia* **14**, 4465–4470 CrossRef Medline

4. Fujita-Yamaguchi, Y., Oishi, K., Suzuki, K., and Imahori, K. (1982) Studies on carbohydrate binding to a lectin purified from Streptomyces sp. *Biochim. Biophys. Acta* **701**, 86–92 CrossRef Medline

5. Harrison, J., and Studholme, D. J. (2014) Recently published *Streptomyces* genome sequences. *Microb. Biotechnol.* **7**, 373–380 CrossRef Medline

6. Hoefler, B. C., Konganti, K., and Straight, P. D. (2013) *De novo* assembly of the *Streptomyces* sp. strain Mgl1 genome using PacBio single-molecule sequencing. *Genome Announc.* 1, e00535-13 CrossRef Medline

7. Fujita, Y. (1976) *Studies on hemagglutinins produced by microorganisms.* Ph.D. thesis, the University of Tokyo, Tokyo, Japan

8. Guerrero, F., Ciragan, A., and Iwai, H. (2015) Tandem SUMO fusion vectors for improving soluble protein expression and purification. *Protein Expr. Purif.* **116**, 42–49 CrossRef Medline

9. Nölling, J., Breton, G., Omelchenko, M. V., Makarova, K. S., Zeng, Q., Nölling, J., and Studholme, D. J. (2014) Recently published *Streptomyces* genome sequences. *Microb. Biotechnol.* **7**, 373–380 CrossRef Medline

10. Sullivan, L., Papoutsakis, E. T., and Bennett, G. N. (2007) Analysis of the clostridial hydrophobic with a conserved trypotypan family (CtW) of proteins in *Clostridium acetobutylicum* with emphasis on ChW14 and ChW16/17. *Biochim. Biophys. Acta* **1578**, 83–94 CrossRef Medline

11. Yasuda, E., Tateno, H., Hirabayashi, J., Ino, T., and Sako, T. (2011) Lectin microarray reveals binding profiles of *Lactobacillus casei* strains in a comprehensive analysis of bacterial cell wall polysaccharides. *Appl. Environ. Microbiol.* **77**, 4539–4546 CrossRef Medline

12. Zhang, Q., Ames, J. M., Smith, R. D., Baynes, J. W., and Metz, T. O. (2009) A perspective on the Maillard reaction and the analysis of protein glycation by mass spectrometry: probing the pathogenesis of chronic disease. *J. Proteome Res.* **8**, 754–769 CrossRef Medline

13. Negre-Salvayre, A., Salvayre, R., Augé, N., Pamplona, R., and Porte-Rothin, M. (2009) Hyperglycemia and glycation in diabetic complications. *Antioxid. Redox Signal.* **11**, 3071–3109 CrossRef Medline

14. Cheng, H. N., and Neiss, T. G. (2012) Solution NMR spectroscopy of food polysaccharides. *Polymer Rev.* **52**, 81–114 CrossRef

15. Fujimoto, Z., Jackson, A., Michikawa, M., Maehara, T., Momma, M., Hennrisat, B., Gilbert, H. J., and Kaneko, S. (2013) The structure of a *Streptomyces avermitilis* α-i–rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement. *J. Biol. Chem.* **288**, 12376–12385 CrossRef Medline

16. Osman, M. E., Williams, P. A., Menzies, A. R., and Phillips, G. O. (1993) Characterization of commercial samples of Gulf arabic. *J. Agricultural Food Chem.* **41**, 71–77 CrossRef

17. Tateno, H., Saneyoshi, A., Ogawa, T., Muramot, K., Kamiya, H., and Saneyoshi, M. (1998) Isolation and characterization of rhamnose-binding lectins from eggs of steelhead trout (*Oncorhynchus mykiss*) homologous to low density lipoprotein receptor superfamily. *J. Biol. Chem.* **273**, 19190–19197 CrossRef Medline

18. Hosono, M., Ishikawa, K., Minek, R., Murayama, K., Numata, C., Ogawa, Y., Takayanagi, Y., and Nitta, K. (1999) Tandem repeat structure of rhamnose-binding lectin from catfish (*Silurus asotus*) eggs. *Biochim. Biophys. Acta* **1472**, 668–675 CrossRef Medline

19. Tateno, H., Ogawa, T., Muramot, K., Kamiya, H., and Saneyoshi, M. (2002) Rhamnose-binding lectins from steelhead trout (*Oncorhynchus mykiss*) eggs recognize bacterial lipopolysaccharides and lipoteichoic acid. *Biosci. Biotechnol. Biochem.* **66**, 604–612 CrossRef Medline

20. Tateno, H., Ogawa, T., Muramot, K., Kamiya, H., and Saneyoshi, M. (2002) Rhamnose-binding lectins from steelhead trout (*Oncorhynchus mykiss*) eggs recognize bacterial lipopolysaccharides and lipoteichoic acid. *Biosci. Biotechnol. Biochem.* **66**, 604–612 CrossRef Medline

21. Terada, T., Watanabe, Y., Tateno, H., Naganuma, T., Ogawa, T., Muramot, K., and Kamiya, H. (2007) Structural characterization of a rhamnose-binding glycoprotein (lectin) from Spanish mackerel (*Scomberomorous niphonius*) eggs. *Biochim. Biophys. Acta* **1770**, 617–629 CrossRef Medline

22. Tateno, H. (2010) SUEL-related lectins, a lectin family widely distributed throughout organisms. *Biosci. Biotechnol. Biochem.* **74**, 1141–1144 CrossRef Medline

23. Ogawa, T., Watanabe, M., Naganuma, T., and Muramot, K. (2011) Diversified carbohydrate-binding lectins from marine resources. *J. Am. Acid.* **2011**, e83914 CrossRef Medline

24. Ozeki, Y., Yokota, Y., Kato, K. H., Titani, K., and Matsui, T. (1995) Developmental expression of D-galactoside-binding lectin in sea urchin (*Asthenosoma crassispina*) eggs. *Exp. Cell Res.* **216**, 318–324 CrossRef Medline

25. Watanabe, Y., Tateno, H., Nakamura-Tsuruta, S., Kominami, J., Hirabayashi, J., Nakamura, O., Watanabe, T., Kamiya, H., Naganuma, T., Ogawa, T., and Naudé, R. J., and Muramot, K. (2009) The function of rhamnose-binding lectin in innate immunity by restricted binding to Gb3. *Dev. Comp. Immunol.* **33**, 187–197 CrossRef Medline

26. Sharon, N. (1987) Bacterial lectins, cell-cell recognition and infectious disease. *FEBS Lett.* **217**, 145–157 CrossRef Medline

27. Ng, S. K., Huang, Y. T., Lee, Y. C., Low, E. L., Chiu, C. H., Chen, S. L., Mao, L. C., and Chang, M. D. (2014) A recombinant horseshoe crab plasma lectin recognizes specific pathogen-associated molecular patterns of bacteria through rhamnose. *PLoS One* **9**, e15296 CrossRef Medline

28. Carpentier, G. (2010) Contribution: protein array analyzer for ImageJ. *ImageJ News* **10**