Differential Recognition and Hydrolysis of Host Carbohydrate Antigens by Streptococcus pneumoniae Family 98 Glycoside Hydrolases

Melanie A. Higgins1,2, Garrett E. Whittworth3, Nahida El Warry1, Mialy Randriantsoa1, Eric Samain1, Robert D. Burke1, David J. Vocadlo4,5, and Alisdair B. Boraston1,6

From 1Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6, Canada, the 2Department of Chemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada, and the 3Centre de Recherches sur les Macromolécules Végétales (CERMAV-CNRS), BP 53, 38041 Grenoble Cedex 9, France

The presence of a fucose utilization operon in the Streptococcus pneumoniae genome and its established importance in virulence indicates a reliance of this bacterium on the harvesting of host fucose-containing glycans. The identities of these glycans, however, and how they are harvested is presently unknown. The biochemical and high resolution x-ray crystallographic analysis of two family 98 glycoside hydrolases (GH98s) from distinctive forms of the fucose utilization operon that originate from different S. pneumoniae strains reveal that one enzyme, the predominant type among pneumococcal isolates, has a unique endo-β-galactosidase activity on the LewisX antigen. Altered active site topography in the other species of GH98 enzyme tune its endo-β-galactosidase activity to the blood group A and B antigens. Despite their different specificities, these enzymes, and by extension all family 98 glycoside hydrolases, use an inverting catalytic mechanism. Many bacterial and viral pathogens exploit host carbohydrate antigens for adherence as a precursor to colonization or infection. However, this is the first evidence of bacterial endoglycosidase enzymes that are known to play a role in virulence and are specific for distinct host carbohydrate antigens. The strain-specific distribution of two distinct types of GH98 enzymes further suggests that S. pneumoniae strains may specialize to exploit host-specific antigens that vary from host to host, a factor that may feature in whether a strain is capable of colonizing a host or establishing an invasive infection.

Streptococcus pneumoniae asymptptomatically colonizes the nasopharynx of 10–40% of people, but given the appropriate opportunity, it can become an extremely aggressive pathogen (1–3). This bacterium causes millions of deaths annually (1), is acquiring antibiotic resistance (4), and shows a disturbing and lethal synergy with the Influenza virus (5). The ability of S. pneumoniae to cause invasive disease is increasingly being linked with the capacity of this bacterium to attack and process the glycans present in host tissues (see Ref. 6 for a review). Indeed, large scale screening of pneumococcal virulence factors has revealed a large complement of genes devoted to complex carbohydrate metabolism that contribute to pneumococcal virulence (7–9). Recent elegant studies have focused on showing how a group of three exo-glycosidases sequentially trim complex human N-glycans (10, 11). These enzymes, however, only make up a fraction of the 39 glycosidases predicted to be in the pneumococcal genome (TIGR4 strain); at least 18 of these 39 are required for full virulence of the bacterium (7). Despite the growing appreciation for the role of carbohydrate metabolism in pneumococcal virulence and the possibility of targeting such metabolic pathways with small molecule therapeutic compounds, the bulk of the carbohydrate-active proteins of S. pneumoniae remain unexamined. As such, we presently have a relatively superficial but growing appreciation for the array of host glycans that S. pneumoniae can degrade.

Several S. pneumoniae genes whose protein products are dedicated to the harvesting and processing of the sugar fucose are beginning to emerge as an important set of pneumococcal virulence factors (12). Comparative genomic studies of several S. pneumoniae genomes has suggested genetic variability at this locus; however, some components of the operon were observed to be present in all of the studied isolates (13). Through our recent identification and characterization of a novel solute-binding protein present in an alternate pneumococcal fucose utilization operon, we have made the observation that there are two different fucose utilization operons distributed among pneumococcal strains (14). Although the organization and composition of the two operons is different, both pathways are predicted to be initiated by the action of a family 98 glycoside hydrolase that is probably secreted (for a discussion of the sequence classification system of glycoside hydrolases, see Ref. 15). This GH98 is the same as that identified as a virulence factor in the TIGR4 strain (7). Remarkably, the GH98 enzymes from the two different pathways display different modular architectures, and their shared catalytic modules only have...
modest amino acid sequence identity. Given the placement of these enzymes in a fucos utilization operon, we hypothesized that they have activity on fucose-containing glycans; however, their divergent sequences and different modular arrangements led us to postulate that they would have different glycansubstrate specificities.

Here we describe the specificity and catalytic mechanism for these two different types of S. pneumoniae GH98 enzymes, one from the TIGR4 strain (Sp4GH98) and the other from the SP3-B571 strain (Sp3GH98). Both enzymes act as endo-β-1,4-galactosidases on the galactosyl-β-1,4-N-acetylglucosamine linkage found in type 2 carbohydrate blood group antigens, although Sp4GH98 displays specificity for the LewisY antigen, whereas Sp3GH98 is highly selective for the same linkage in the blood group A/B-antigens. The biochemical analysis of these enzymes in combination with the determination of their structures in complex with products and substrates provides molecular level insight to their catalytic mechanism and how they discriminate between their respective substrates. We discuss these results in the context of the recent association of the pneumococcal fucose utilization operon with the virulence of S. pneumoniae (7, 12) and the possible strain-specific dependence of pneumococcal virulence on the carbohydrate antigens presented by different hosts.

**EXPERIMENTAL PROCEDURES**

**Materials**—Sodium hydroxide (50% solution) was purchased from Fisher. D-Glucose was purchased from Sigma, the LewisY tetrasaccharide was purchased from V-LABS, Inc. (Covington, LA), and the blood group A and B trisaccharides were purchased from Dextra Laboratories (Reading, UK). Milli-Q (18.2 megaohms cm−1) water was used to prepare all eluents, buffers, and standards, whereas a 4× phosphate buffer solution (4× PBS)5 was made using salts purchased from Bioshop (Burlington, Canada). A-LewisY was purchased from Dextra Laboratories. The A and B tetrasaccharides used to obtain the Sp3GH98 structures were obtained from Core D of the Consortium for Functional Genomics. All other sugars were purchased in metabolically engineered Escherichia coli strains.6

**Cloning**—The catalytic module encoding gene fragments from Sp4GH98 and Sp3GH98 were amplified by PCR from S. pneumoniae TIGR4 genomic DNA (American Type Culture Collection BAA-334D) and S. pneumoniae SP3-B571 genomic DNA (kindly provided by Dr. Garth Ehrlich), respectively, using primers to introduce 5′ Nhel and 3′ Xhol restriction endonuclease sites (supplemental Table S5). A “megaprimer” PCR method was employed to introduce site-directed amino acid substitutions (16). All amplified DNA fragments were cloned into pET28a (Novagen) via the engineered Nhel and Xhol restriction sites using standard molecular biology procedures. The DNA sequences of these constructs were verified by bidirectional sequencing.

**Protein Production and Purification**—Both Sp4GH98 and Sp3GH98 were produced in E. coli BL21 Star (DE3) cells using LB medium supplemented with 50 μg ml−1 of kanamycin. Cultures were grown at 37 °C to an optical density at 600 nm of 0.5–0.7, induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside, and further incubated at 16 °C overnight. Cells were harvested by centrifugation and ruptured by chemical lysis. Target proteins were purified by Ni²⁺ immobilized metal affinity chromatography followed by size exclusion chromatography using a Sephacryl S-200 column (GE Biosciences). Size exclusion chromatography was performed using 20 mM Tris-HCl, pH 8.0, and 250 mM NaCl as the buffer for Sp3GH98 and 20 mM Tris-HCl, pH 8.0, and 1 mM dithiothreitol for Sp4GH98. Selenomethionine-labeled Sp4GH98 was prepared using previously described procedures (17) and purified as above.

Protein concentration was determined by measuring the absorbance at 280 nm and using calculated molar extinction coefficients of 0.156440 cm−1 μM−1 for Sp4GH98 and 0.142670 cm−1 μM−1 for Sp3GH98 (18).

**NMR Analysis**—¹H NMR spectroscopy (600-MHz Bruker AMX spectrometer) was used to follow the progress and identify the products of the Sp4GH98-catalyzed reaction. The reaction was carried out in ~0.6 ml of PBS, pH 7.4, made up in D₂O containing 2.7 mM LewisY tetrasaccharide. The reaction was initiated by the addition of 15 μl of a 35 mg/ml stock of the Sp4GH98. The hydrolysis of the LewisY tetrasaccharide was monitored until the reaction reached equilibrium. An initial spectrum (referred to as time 0) containing substrate and buffer was acquired before the addition of enzyme, whereas the product spectrum containing the H disaccharide and buffer was acquired with no addition of enzyme.

**Enzyme Assays**—Carbohydrates were separated by high performance anion exchange chromatography with pulsed amperometric detection (HPAE-PAD) using a Dionex ICS 3000 HPLC equipped with ASI 100 Automated sample injector (Dionex) and an ED₅₀ electrochemical detector (Dionex) with a gold working electrode and an Ag/AgCl reference electrode.

All assays were carried out in duplicate at 37 °C for 30 min using a stopped assay procedure. A 50-μl assay volume was used for all standard solutions with 12.5 μl of 4× PBS (200 mM NaH₂PO₄ and 400 mM sodium chloride) buffering the solution to pH 7.4. Assays were initiated by the addition of enzyme (2.5 μl). Upon the addition of 200 μl of anhydrous ethanol to halt the reaction, 10 μl of an internal standard D-glucose (from a 1 mM stock solution) was added to give a 40 μM final concentration of D-glucose, and the assay was stored at −20 °C for 20 min. Precipitated protein was removed by centrifugation at 14,000 rpm and 4 °C for 5 min in a microcentrifuge. The supernatant containing mono- and oligosaccharides was carefully removed from each tube, transferred to a new tube, and evaporated to dryness in a vacuum centrifuge. All samples were reconstituted in 250 μl of water. A 20-μl injection volume of samples was separated with a PA-100 column set (analytical plus guard column) and a 100 mM NaOH isocratic gradient for 20 min.

Standard calibration curves were generated for the LewisY tetrasaccharide, blood group A type 2 and B type 2 pentasaccharides, H antigen disaccharide, and the blood group A and B trisaccharides. The standard calibration curves for each substrate were analyzed at pH 7.4 and ranged from 25 to 500 μM,
Hydrolysis of Host Carbohydrate Antigens

whereas a similar curve for the monitored products ranged from 7 to 500 μM.

Substrate specificity assays of Sp3GH98 (18 mg/ml) and Sp4GH98 (34 mg/ml) were carried out using the Lewis^3 tetrasaccharide and blood group A and B type 1, 2, and 4 pentasaccharides. Each substrate was individually incubated at concentrations of 100 and 500 μM with Sp3GH98 and Sp4GH98 at 37 °C for 30 min, using the aforementioned stopped assay procedure.

Time-dependent assays of Sp3GH98 and Sp4GH98 revealed that both enzymes were stable over a 5-min period in the PBS buffer at pH 7.4. A range of 0.05–2 mM substrate Lewis^3 was incubated with Sp4GH98 (0.17 mg/ml) for 5 min at 37 °C, and the assay was halted, concentrated, and reconstituted as described above. For the blood group A and B type 2 pentasaccharide substrates, a range from 0.025 to 5 mM was incubated with Sp3GH98 (0.0045 mg/ml) for 5 min at 37 °C, and the assay was halted, concentrated, and reconstituted as described above. Separation and detection followed the previously mentioned strategy, and quantification of the substrates and products was calculated relative to the calibration curves.

Immunofluorescence Microscopy—Cells (A549, ATCC CCL-185) were cultured in Dulbecco’s modified Eagle’s medium/F-12, containing 10% fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, and passaged into 8-well slide chambers. Cells were rinsed in PBS and fixed in 4% paraformaldehyde in PBS (15 min), rinsed, and blocked in 5% lamb serum in PBS/Tween 20 (15 min, room temperature). Cells were treated with 2.5 mg/ml enzyme (Sp3GH98 and/or Sp4GH98), diluted in 20 mM Tris-HCl, pH 8.0, with 1 mM dithiothreitol for Sp4GH98 or 250 mM NaCl for the Sp3GH98 (overnight, room temperature). Cells were rinsed three times in PBS before application of primary antibodies. Antibodies used were anti-Lewis^a antigen (Abcam; ab23911-100) and anti-A/B antigen (Abcam; ab24223). Primary antibodies were diluted 1:50 (Lewis^3) and 1:400 (A/B) in 5% lamb serum in PBS and applied to the fixed cells overnight (4 °C). Specimens were rinsed three times in PBS, and Alexa 568-conjugated goat anti-mouse IgG secondary antibody (Molecular Probes) diluted 1:800 in PBS was applied.

After incubation (1–2 h), cells were rinsed with PBS and counterstained with Hoechst 33342 (1 ng/ml; Sigma). Slides were mounted with coverslips and examined with a Leica DM-6000 epifluorescence microscope, and images were captured with a Hamamatsu Orca wide field camera controlled with Openlab software (version 4.04). Overall contrast and brightness were adjusted, and images were cropped and assembled using Photoshop (CS2). For quantification, fluorescence intensity per unit area was determined within a 5 x 5-μm sample area placed over the cytoplasm of cells chosen at random in monochrome images using ImageJ (version 1.40g). Statistical analysis (one-way analysis of variance, Bonferroni’s multiple comparison test) was done with GraphPad Prism (version 4.03).

Crystallization—All crystals were obtained by hanging drop vapor diffusion at 18 °C. In all cases, final diffraction quality crystals were obtained when 2–4 μl of protein was incubated with the same volume of crystallization solution at room temperature for 5–10 min and then centrifuged for 10 min at 13,000 rpm prior to setting up the crystallization experiment. The highest quality crystals were grown when seeded.

Crystals of Selenomethionine-Sp4GH98 (20 mg/ml) were obtained in 19% (v/v) polyethylene glycol 4000, 0.2 M sodium acetate, 0.1 M trisodium citrate, pH 5.6, and 3 mM dithiothreitol. Native Sp4GH98 crystals and mutant Sp4GH98 E158A crystals were obtained using protein at 20 mg/ml in 19% (v/v) polyethylene glycol 4000, 0.2 M sodium formate, 0.1 M trisodium citrate, pH 5.6, and 2 mM dithiothreitol. Native Sp4GH98 crystals were soaked in the crystallization solution containing a molar excess of Lewis^3 tetrasaccharide for 45 min to yield H disaccharide product complex, whereas mutant Sp4GH98 E158A crystals were soaked in the crystallization solution supplemented with molar excess Lewis^a pentasaccharide for 20 min to produce a substrate complex.

Native Sp3GH98 and mutant Sp3GH98 E558A, both at 20 mg/ml, were crystallized in 17% (v/v) polyethylene glycol 3350, 0.2 M ammonium sulfate, and 0.1 M sodium acetate trihydrate, pH 4.8. Native crystals were soaked in the crystallization solution supplemented with a molar excess of A or B blood group antigen tetrasaccharides for 45 min to generate A and B blood group antigen trisaccharide product complexes, respectively, whereas crystals of Sp3GH98 E558A were soaked with A-Lewis^a antigen pentasaccharide for 20 min to produce a substrate complex.

Data Collection, Structure Determination, and Refinement—Crystals were flash-cooled with liquid nitrogen in crystallization solution supplemented with 20–30% (v/v) ethylene glycol. Data were processed using Crystal Clear/d*trek (19) or MOSFLM/SCALA (20, 21). All data collection and processing statistics are shown in Tables 1 and 2.

The structure of the selenomethionine-Sp4GH98 structure was solved by a single anomalous dispersion experiment using a x-ray wavelength optimized for the f^3’ of selenium (determined by a fluorescence scan). The positions of 16 of the 18 selenium atoms expected for the single Sp4GH98 molecule in the asymmetric unit were determined using Shelx/C/D (22) with data extending to 3.0 Å. Initial phases were produced by refinement of the selenium substructure parameters with SHARP (23) using data to 2.2 Å followed by phase improvement and extension to 1.6 Å with DM (24). Using the phase output from DM, ARP/wARP (25) was able to build a nearly complete model of Sp4GH98 with docked side chains. This initial model was used as a starting point for the building and refinement of Sp4GH98 using the 1.5 Å resolution native data set. The model was completed using COOT (26), followed by refinement using REFMAC (27). Due to the collection of the native data on a square detector, the high resolution reflections present in the corners of the detector could not be collected to completeness. Although incomplete, the high resolution data were nevertheless judged on the basis of the Rmerge and I/σI of the high resolution bin to be of sufficiently high quality to retain in the data set.

The native Sp4GH98 coordinates were used to solve the structure of Sp3GH98 in complex with the A blood group antigen trisaccharide by molecular replacement. A molecular replacement solution comprising the two Sp3GH98 molecules in the asymmetric unit was found with PHASER (28). This ini-
**Hydrolysis of Host Carbohydrate Antigens**

**TABLE 1**

Data collection and refinement statistics for Sp4GH98  
Values in parentheses are for the highest resolution shell. NA, not available.

| Data collection          | Sp4GH98 | Sp4GH98 + product | Sp4GH98E158A + Lewis^* |
|-------------------------|---------|------------------|------------------------|
| Beamline                | NSLS X-C | SSRL 7.1         | CLS CMCF1              |
| Wavelength              | 0.9794 Å | 0.9761 Å         | 0.9214 Å               |
| Space group             | P2 2,2, 2 | P2 2,2, 2        | P2 2,2, 2              |
| Cell dimensions: a, b, c (Å) | 50.8, 91.7, 116.0 | 50.8, 90.5, 116.8 | 50.8, 90.8, 116.5 |
| Resolution (Å)          | 20.00-1.60 (1.66-1.60) | 20.00-1.50 (1.58-1.50) | 20.00-1.70 (1.76-1.70) |
| Rmerge (%)              | 0.118 (0.410) | 0.060 (0.302) | 0.127 (0.413) |
| I/σ (r)                 | 20.3 (4.2) | 16.9 (3.6) | 11.0 (4.1) |
| Completeness (%)        | 99.9 (99.9) | 97.4 (96.1) | 97.7 (97.7) |
| Redundancy              | 6.8 (6.7) | 3.9 (3.6) | 4.8 (4.6) |
| Refinement              |         |                  |                        |
| Resolution (Å)          | 1.50    | 1.70             | 2.28                   |
| No. of reflections      | 69668   | 55648            | 22735                  |
| Rmerge/Rfree            | 0.13/0.16 | 0.15/0.20 | 0.16/0.23              |
| No. of atoms            |         |                  |                        |
| Protein                 | 4503    | 4490             | 4498                   |
| Ligand                  | NA      | 22               | 46                     |
| Water                   | 600     | 642              | 334                    |
| β-factors               |         |                  |                        |
| Protein                 | 12.6    | 10.3             | 12.5                   |
| Ligand                  | N/A     | 19.9             | 31.4                   |
| Water                   | 24.1    | 22.4             | 16.4                   |
| Root mean square deviations |       |                  |                        |
| Bond lengths (Å)        | 0.016   | 0.015            | 0.010                  |
| Bond angles (degrees)   | 1.597   | 1.524            | 1.304                  |
| Ramachandran            |         |                  |                        |
| Preferred (%)           | 99.2    | 99.0             | 99.4                   |
| Generously allowed (%)  | 0.4     | 0.6              | 0.2                    |
| Disallowed (%)          | 0.4     | 0.4              | 0.4                    |

Results and Discussion

*S. pneumoniae* GH98 Enzymes—The GH98 enzymes from *S. pneumoniae* TIGR4 (Sp4GH98) and SP3-BS71 (Sp3GH98) are 1038- and 1005-amino acid, respectively, multimodular proteins with predicted classical N-terminal Gram-positive secretion signal sequences. Sp4GH98 possesses three C-terminal family 47 carbohydrate-binding modules (33), whereas Sp3GH98 has two N-terminal modules that have sequence identity with family 51 carbohydrate-binding modules (34). The N terminus of Sp4GH98 and the C terminus of Sp3GH98 contain a domain of ~650 amino acids that is predicted on the basis of sequence alignments with a known family 98 glycoside hydrolase, the GH98 endogalactosidase E-ABase from *Clostridium perfringens* (35), to house the catalytic activity. An amino acid sequence-based comparison of the catalytic modules reveals that Sp4GH98 shows 34% amino acid identity with *C. perfringens* E-ABase, whereas Sp3GH98 shows 60% amino acid sequence identity with E-ABase. The catalytic modules of Sp4GH98 and Sp3GH98 show only 35% amino acid identity with each other. Genes with 99% or greater DNA sequence identity to that encoding Sp4GH98 are found in 19 of the 23 sequenced genomes of *S. pneumoniae*; the remaining four strains have genes encoding Sp4GH98-type enzymes (see supplemental Table 1). Although the sample size is relatively small, the distribution of the two types of GH98 enzymes among pneumococcal strains appears to be independent of strain serotype (see supplemental Table S1).

To facilitate the biochemical and structural analysis of Sp4GH98 and Sp3GH98, we used our amino acid sequence alignments to precisely define the catalytic domain in these enzymes. Guided by this information, we cloned the DNA fragments encoding only the catalytic domains and overproduced the proteins in *E. coli*. For simplicity, we will continue to refer to these truncated constructs as Sp4GH98 and Sp3GH98.

GH98 Enzymes Are Active on Blood Group Antigens in Vitro and in Situ—Given the different modular structures of Sp3GH98 and Sp4GH98 and the presence of the genes encoding these divergent GH98 enzymes within fucose utilization operons having different gene contents and organizations (14), we suspected that these enzymes might have different, yet related, substrate specificities for human cell surface fucose-containing glycans. Furthermore, previous studies found that E-ABase cleaved the A and B blood group glycotopes from protein substrates and cell surfaces (35). Accordingly, we set out to evaluate the substrate specificity of these two *S. pneumoniae* enzymes against select fucose-containing blood group antigens,
including pentasaccharides of blood group glycotopes A and B containing type 1, 2, and 4 core chains, the type 2 H trisaccharide, the LewisY antigen, and the Lewisb antigen (see supplemental Fig. S1 for a summary of the structures of glycans used in this study).

To assess the activity of Sp3GH98 and Sp4GH98 against these oligosaccharides, we used HPAE-PAD (see supplemental Fig. S1 for a summary of these data). Of these substrates, Sp4GH98 cleaved only the LewisY antigen structures to liberate the A antigen (Fucα1→2Gal) from the remaining portion of the glycan, suggesting that the active site of this enzyme cannot accommodate the terminal GalNAc or Gal moiety of the A and B blood group antigens. The catalytic specificity of Sp4GH98 for the LewisY antigen is consistent with the known preference of the enzyme’s three C-terminal family 47 carbohydrate-binding modules for this antigen (33). Furthermore, results that are publicly available through the Consortium for Functional Glycomics regarding the screening of complete, intact Sp4GH98 (referred to as “fucolectin-related protein”) against a glycan microarray also revealed the LewisY antigen to be the single major glycan with which this protein interacts. In contrast, we found that Sp3GH98 did not cleave the LewisY tetrasaccharide but did process the pentasaccharide blood group glycotopes A and B, depending, however, on the structure of the core chain. In keeping with its high amino acid sequence identity with C. perfringens E-Base, we found that Sp3GH98 is specific for the Galβ1→4GlcNAc (type 2 core chain) linkage over either the Galβ1→3GlcNAc (type 1 core chain) linkage or the Galβ1→3GalNAc (type 4 core chain) and liberates the terminal GalNAcα1→3(Fucα1→2)Gal trisaccharide of the A antigen and the terminal Galα1→3(Fucα1→2)Gal trisaccharide of the B antigen.

### Table 2
Data collection and refinement statistics for Sp3GH98

| Data collection | Sp3GH98+A-trisaccharide | Sp3GH98+B-trisaccharide | Sp3GH98E558A+A-pentasaccharide |
|----------------|-------------------------|-------------------------|---------------------------------|
| Beamsline      | SSR1 7.1                | SSR1 7.1                | CLS CMCF1                       |
| Wavelength     | 0.9761 Å                | 0.9761 Å                | 0.9793 Å                        |
| Space group    | P2, 2, 2                | P2, 2, 2                | P2, 2, 2                        |
| Cell dimensions: a, b, c (Å) | 97.7, 153.8, 91.0  | 98.1, 154.2, 91.2  | 97.9, 153.7, 90.9  |
| Resolution (Å) | 40.00-1.90 (2.00-1.90) | 40.00-2.00 (2.11-2.00) | 40.00-1.90 (2.00-1.90) |
| Rmerge         | 0.087 (0.350)           | 0.063 (0.224)           | 0.113 (0.368)                   |
| Completeness (%) | 97.0 (98.1)            | 84.6 (74.0)            | 99.1 (98.0)                     |
| Redundancy     | 3.6 (3.7)               | 6.1 (5.4)               | 6.7 (6.4)                       |
| Refinement     |                         |                         |                                 |
| Resolution (Å) | 1.90                    | 2.00                    | 1.90                            |
| No. of reflections | 99648                | 74917                   | 102091                          |
| Rwork/Rfree   | 0.17/0.21               | 0.16/0.21               | 0.17/0.21                       |
| No. of atoms   |                         |                         |                                 |
| Protein        | 4681 (A); 4670 (B)      | 4681 (A); 4673 (B)      | 4677 (A); 4682 (B)              |
| Ligand         | 36 (C); 36 (D)          | 33 (C); 33 (D)          | 60 (C); 60 (D)                  |
| Water          | 1024                    | 981                     | 974                             |
| Root mean square deviations |             |                         |                                 |
| Bond lengths (Å) | 0.018                 | 0.016                   | 0.019                           |
| Bond angles (degrees) | 1.588                  | 1.474                   | 1.666                           |
| Ramachandran   |                         |                         |                                 |
| Preferred (%)  | 98.5                    | 98.4                    | 98.4                            |
| Generally allowed (%) | 0.9                   | 0.8                     | 0.8                             |
| Disallowed (%) | 0.6                     | 0.8                     | 0.8                             |

### Table 3
Kinetic analysis of GH98 activity

| Enzyme | Substrate | kcat | Km | kcat/Km |
|--------|-----------|------|----|--------|
| Sp4GH98 | LewisY | ND | ND | 0.026 ± 0.001 |
| Sp3GH98 | A antigen | 280 (±10) | 440 (±40) | 64 ± 10 |
|         | B antigen | 670 (±20) | 570 (±60) | 1.2 ± 0.1 |

*ND, not determined.

To quantitatively evaluate the processing of these substrates by Sp3GH98 and Sp4GH98, we carried out enzyme kinetics studies following the reaction by using HPAE-PAD where the production of the A or B trisaccharides (for Sp3GH98) or the H disaccharide (for Sp4GH98) was monitored in conjunction with an internal standard (see supplemental Fig. 2). Time-dependent assays revealed that both enzymes were stable at 37 °C over the 5-min assay period and that, despite consumption of greater than 10% of substrate, the rate of disappearance of substrate and the formation of product were both linear over this period (supplemental Fig. S3), enabling this experimental approach for evaluating the kinetics of these two enzymes. Therefore, by varying the substrate concentrations, we were able to establish kinetic parameters governing the enzyme-catalyzed hydrolysis of these substrates (Table 3). For the Sp3GH98-catalyzed hydrolysis of the A and B antigens, we found clear saturation kinetics (Fig. 1) yielding very comparable kinetic parameters (Table 3). Interestingly, for the Sp4GH98-catalyzed hydrolysis of the LewisY tetrasaccharide, we did not observe substrate saturation (Fig. 1), making it possible only to determine the second order rate constant, kcat/Km (Table 3). The absence of saturation even at concentrations of 2 mM LewisY tetrasaccharide suggests either that the enzyme may...
have additional binding subsites in its active site that are not exploited by the tetrasaccharide or, alternatively, that the effective concentration of substrate in the natural environment is much higher when the enzyme is bound to the cell surface. The rationale we favor is that the carbohydrate-binding modules present in the intact Sp4GH98 and Sp3GH98 enzymes probably function in vivo to target these enzymes to their respective tissue-presented carbohydrate substrates, localizing them and maintaining their proximity to their glycan substrates, exactly as carbohydrate-binding modules have been shown to do in plant cell wall-degrading glycoside hydrolases (36). This effect would enable the intact enzymes to overcome some of the limitations inherent in the apparently high $K_m$ values of the catalytic modules.

With knowledge of the in vitro activity of Sp4GH98 and Sp3GH98, we set out to determine if these proteins have activity on cell surface-presented substrates. Given that the tissue most commonly targeted by S. pneumoniae is in the lungs, we approached this question by using a type II alveolar cell line (A549), which is commonly used to assess pneumococcal adherence and invasiveness, with the presumption that treatment of the cell line with one or both enzymes should remove the carbohydrate antigen from the cells in a manner consistent with the specificity of these enzymes.

Cells not treated with enzymes and probed with the anti-Lewis$^y$ antibody had small foci of fluorescence scattered over the surface of the cells (Fig. 2A). Notably, cells treated with Sp4GH98 and probed with the anti-Lewis$^y$ antibody had a significant reduction in the number and intensity of the immunoreactive foci, consistent with the Sp4GH98-catalyzed destruction of Lewis$^y$ antigen on the cells (Fig. 2, B and I). In contrast, when cells were treated with Sp3GH98 and probed with the anti-Lewis$^y$ antibody, the fluorescent foci were smaller and more dispersed, but the mean fluorescence was not significantly reduced (Fig. 2, C and I). When cells were digested with both enzymes, anti-Lewis$^y$ immunoreactivity was again reduced significantly but not more so than when treated with Sp4GH98 alone (Fig. 2, D and I). In a parallel set of experiments, in which cells were probed after enzyme treatment with an antibody that recognizes both the A and B blood group antigens, only pretreatment of the cells with Sp3GH98 alone or with both enzymes resulted in significant decreases in fluorescence intensity (Fig. 2, E–I). The fairly high concentrations of enzyme ($\sim 40 \mu M$) required to observe significant decreases in cell surface glycosylation are consistent with the relatively high $K_m$ ($\sim 400 \mu M$) values observed for these enzymes and are likely a consequence of the use of the recombinant proteins having the catalytic domains but lacking the carbohydrate-binding modules. Regardless, the processing of cell surface A and B blood group antigens by Sp3GH98 is also consistent with studies showing that the C. perfringens E-ABase processes these antigens on erythrocytes (35). Collectively, these results clearly indicate that both S. pneumoniae Sp4GH98 and Sp3GH98 are capable of removing carbohydrate antigens from a model lung cell line and with activity in keeping with their in vitro specificity.

**Structure of GH98 and Specific Glycon Recognition**—A small number of glycoside hydrolases are now known to be active on blood group antigens (35, 37, 38). Other than E-ABase, these are exoglycosidases that remove the non-reducing terminal A or B antigen-determining N-acetylgalactosamine or galactose residues, respectively, converting the antigen to the H-type (O-type) (37). The specificity of Sp3GH98 is thus quite different and akin to that of C. perfringens E-ABase, whereas the activity of Sp4GH98 on the Lewis$^y$ antigen has not yet been observed in any enzyme. This difference and the lack of a structure of any

FIGURE 1. Kinetic properties of S. pneumoniae family 98 enzymes. Shown are kinetic plots of the activity of Sp3GH98 on the type 2 A pentasaccharide (A) and the type 2 B pentasaccharide (B), C, a corresponding analysis of the activity of Sp4GH98 on the Lewis$^y$ tetrasaccharide.
member of this family of GH98 enzymes prompted us to explore the structural basis of their unique substrate specificity using x-ray crystallography. The x-ray crystal structure of Sp4GH98 was determined to 1.6 Å resolution by a single wavelength anomalous dispersion experiment optimized for selenium. The resulting initial model was subsequently used to complete and refine a native structure to 1.5 Å resolution. The N-terminal domain of the crystallized construct comprises a classical \((\alpha/\beta)_8\)-barrel (Fig. 3A). This module is followed by an 11-stranded \(\beta\)-sandwich domain. An \(\alpha\)-helical insertion in the \(\beta\)-sandwich domain extends out and packs against the \((\alpha/\beta)_8\) barrel. Likewise, a small insertion between \(\beta\)-strand 7 and \(\alpha\)-helix 8 of the \((\alpha/\beta)_8\) barrel interacts with the extension from the \(\beta\)-sandwich. The substantial interactions between the \((\alpha/\beta)_8\) barrel and the \(\beta\)-sandwich create what appears to be a relatively rigid structure. The structure of Sp3GH98 is very similar and overlaps Sp4GH98 with a root mean square deviation of 1.4 Å over 511 matched C atoms (determined by secondary structure matching (39)) (data not shown). Indeed, the relative placements of the constituent domains are nearly identical for the two enzymes, suggesting that they function as a single, rigid structural unit.

To provide insight into the substrate specificities of Sp4GH98 and Sp3GH98, complexes of these enzymes with their reaction products were obtained by soaking crystals of catalytically active proteins with their respective substrates. The structure of Sp4GH98 incubated with the Lewis\(^Y\) antigen revealed unmistakable electron density for a disaccharide in the active site. This electron density could be easily modeled as the H disaccharide (Fuc\(^1\)–2Gal) but not as any other component of the Lewis\(^Y\) antigen (Fig. 3B). Likewise, the structures of Sp3GH98 incubated with the blood group A and B tetrasaccharides yielded unambiguous electron density for the blood group A and B trisaccharides, respectively, in the active site (Fig. 3C; only the A trisaccharide complex is shown as representative data). These results are consistent with the product analysis obtained using HPAE and provide additional supporting evidence that the Gal\(^1\)–4GlcNAc linkage common to carbohydrate antigens having the core 2 chain is hydrolyzed by these GH98 enzymes.

The galactosyl residue of the H disaccharide product in the active site of Sp4GH98 occupies what must be the \(-1\) subsite, whereas the \(\alpha\)-1,2-linked fucosyl residue occupies the \(-2\) subsite (Fig. 3B; subsite nomenclature as in Ref. 40). The chemical groups of this disaccharide make a number of hydrogen bonding and van der Waals interactions with the active site (Fig. 3D).

---

**FIGURE 2. GH98 enzymes remove carbohydrate antigens from a model cell line.** Shown are epifluorescent images of A549 cells treated with Sp4GH98 and Sp3GH98 enzymes and then prepared with antibodies to the Lewis\(^Y\) antigen (red) or the A/B antigens (red) and counterstained with 4,6-diamidino-2-phenylindole (blue). A–D, cells were probed with anti-Lewis\(^Y\) antibody after no treatment (A), Sp4GH98 treatment (B), Sp3GH98 treatment (C), or treatment with both enzymes (D). E–H, cells were probed with anti-A/B antibody after no treatment (E), Sp4GH98 treatment (F), Sp3GH98 treatment (G), or treatment with both enzymes (H). Bar, 5 μm. I, quantification of fluorescence of A549 cells treated as in A–H (treatment is indicated below the graph). Bars, mean ± S.E. fluorescence per unit area for a sampling of five randomly selected regions of cytoplasm. Bars marked with an asterisk, significantly different from cells not treated with an enzyme as judged by one-way analysis of variance.
FIGURE 3. The structural features of *S. pneumoniae* family 98 enzymes. A, divergent stereo schematic representation of the structure of Sp4GH98. The N-terminal \(\alpha/\beta\) barrel housing the catalytic residues is shown in green, the C-terminal \(\beta\)-sandwich domain in purple, and the smaller intervening domain in yellow. Shown are solvent-accessible surface representations (gray) of the active site of Sp4GH98 containing the H disaccharide (shown in a stick representation) (B) and Sp3GH98 containing the A trisaccharide (shown in a stick representation) (C). Green, galactose; blue, fucose; magenta, N-acetylgalactosamine. The yellow mesh shows the maximum likelihood \(2F_o-2F_c\) maps for the carbohydrates (contoured at 3σ; 0.24 electrons/Å\(^3\) for Sp4GH98 and 0.20 electrons/Å\(^3\) for Sp3GH98). Subsites of the active site are labeled in white. D, schematic of the interactions in the \(-1\) and \(-2\) subsites of Sp4GH98 and Sp3GH98. Residues in black are conserved in both enzymes and labeled in black (Sp4GH98) or gray (Sp3GH98). Red residues are present only in Sp4GH98. E, schematic of the \(-2^\prime\) subsite of Sp3GH98. Black residues are shown for interactions that are conserved for the galactose of the B-antigen and the N-acetylgalactosamine of the A antigen. The red residue represents an interaction that is unique to the A antigen. F, divergent stereo representation of a structural overlay of the active sites of Sp4GH98 (blue) and Sp3GH98 (orange). The A trisaccharide product in the Sp3GH98 active site is shown in sticks and colored as in C. The loop protruding from the \(\beta\)-sandwich domain is shown in a schematic diagram representation. Trp\(^{512}\) of Sp4GH98 that blocks the \(-2^\prime\) subsite in this protein is shown in a stick representation.
**Hydrolysis of Host Carbohydrate Antigens**

The active site of Sp3GH98 also accommodates the H-determining Fucα1–2Gal motif in −1 and −2 subsites that are structurally very similar to those of Sp4GH98 (Fig. 3D). The most striking difference between the active sites of Sp3GH98 and Sp4GH98 is the presence of an additional subsite found in Sp3GH98, which enables binding of the branched blood group antigen A or B trisaccharide antigens (Fig. 3C and E). This subsite accommodates the α-1,3-linked terminal N-acetylgalactosamine (Group A) or galactose (Group B) residue. A comparison of the accessible surfaces of the two enzymes’ active sites reveals that the additional Sp3GH98 subsite, which we term the −2′ subsite, is completely absent in Sp4GH98 (Fig. 3, compare B and C). This subsite is largely conserved between Sp4GH98 and Sp3GH98; dele-
tion of the carboxylate of the general acid catalyst has been shown, almost invariably, to have deleterious effects on the glycoside hydrolase-mediated processing of substrates having a carbohydrate leaving group (42). Generation of these mutant proteins resulted in enzymes that crystallized yet had sufficiently low activity in the crystalline state so as to enable us to determine the structures of Sp4GH98E158A and Sp3GH98E558A in complex with intact LewisY and blood group A-LewisY antigen substrates, respectively. Clear electron density for both saccharides allowed the substrates to be readily modeled (Fig. 4, A and B). The structures of the substrate complexes revealed the structural details of two additional subsites, +1 and +1’. The +1 subsite in both enzymes is conserved, and a tryptophan residue (Trp161 in Sp4GH98 and Trp561 in Sp3GH98) that interacts with the GlcNAc residue of the substrate plays a key role by forming a classic protein-carbohydrate interaction whereby the pyranose ring lies parallel to the plane of the indole ring (Fig. 4C). The +1’ subsite accommodates the α-1,3-linked fucose of the LewisY antigen of both potential substrates, LewisY and blood group A-LewisY. This subsite in Sp4GH98 snugly houses this fucose residue and makes numerous van der Waals interactions (Fig. 4A); in particular, the C6-methyl group of the fucose fits into a hydrophobic pocket formed by the apolar side chains of Trp161, Ala130, Ile561, and Thr95 (not shown). This close fit is further complemented by both direct and water-mediated hydrogen bonds (Fig. 4D). Likewise, the blood group A-LewisY antigen complex of Sp3GH98 reveals that the α-1,3-linked fucose residue is comfortably accommodated in the +1’ subsite of this enzyme; however, this subsite is particularly spacious in Sp3GH98 (Fig. 4B), making limited van der Waals interactions and no direct hydrogen bonds with the fucose residue (Fig. 4D). Sp3GH98 clearly does not require the α-1,3-linked fucose residue as a specificity determinant; however, we speculate that the spacious +1’ subsite in Sp3GH98 allows it to also act on the A/B-LewisY antigens by loosely accommodating the α-1,3-linked fucose. For Sp4GH98, however, the secure fit of the LewisY-determining fucose residue provides a rationale for why the α-1,3-linked fucose in the substrate aglycon plays an important role in substrate recognition or catalysis in this enzyme.

In summary, these two subtypes of GH98 enzymes recognize the type 2 H trisaccharide core of their respective substrates through very similar interactions at the +1, −1, and −2 subsites, whereas unique structural aspects of the −2’ and +1’ subsites impart divergent substrate specificity to these two enzymes. In the Sp3GH98, the −2’ subsite accommodates the terminal GalNAc and Gal residues that define the blood group A and B antigens, respectively. This subsite is occluded in Sp4GH98, however, preventing binding of the A/B-glycotopes. Instead, in Sp4GH98, the +1’ subsite specifically interacts with an α-1,3-linked fucose, a residue that is a defining feature of the LewisY antigen. In contrast, Sp3GH98 can accommodate this α-1,3-linked fucose residue but, because of an altered +1’ subsite, makes limited interactions with it and does not require this residue to be present in its substrate for activity.

FIGURE 5. NMR analysis of GH98 catalytic mechanism. A, structure of the LewisY tetrasaccharide substrate. B, structure of the Fuc1–2Gal (H disaccharide) product. C, the hydrolysis of LewisY tetrasaccharide measured as a function of time by 1H NMR spectroscopy. Peaks corresponding to the chemical shifts of the proton on the anomeric carbon for substrate (S) and product (P) are labeled for the α- and β-anomers.
Family 98 Glycoside Hydrolases Use an Inverting Catalytic Mechanism—One defining feature of glycoside hydrolases is the general catalytic mechanism that is used (42). Many families of glycoside hydrolases use a retaining catalytic mechanism in which the substrate is cleaved to generate a hemiacetal product with retained stereochemistry at the anomeric center. Other enzymes use an inverting mechanism, which, as the name implies, results in a product having inverted stereochemistry at the anomeric center. Knowledge of the catalytic mechanism is useful for engineering these enzymes for biotechnology (43) and to design effective inhibitors (44, 45). To assess the general catalytic mechanism used by GH98 enzymes, we carried out analysis of the stereochemical outcome of the reaction catalyzed by Sp4GH98 using 1H NMR. A time course of the hydrolysis of the LewisY tetrasaccharide (Fig. 5A) reveals production of the \(\alpha\)-hemiacetal of the H disaccharide (Fig. 5B) precedes that of the \(\beta\)-hemiacetal (Fig. 5C), indicating hydrolysis of the \(\beta\)-glycosidic linkage via an inverting mechanism. This observation is entirely consistent with the structural features of the active sites of Sp4GH98 and Sp3GH98, which have identical arrangements of conserved active site residues in the \(\alpha\)-1 subsite (Fig. 4C). The side chains of Glu\(^{710}\) and Asp\(^{657}\) in Sp3GH98 and Glu\(^{301}\) and Asp\(^{251}\) in Sp4GH98 coordinate a water molecule that lies \(3.1\) Å directly beneath the C1 of the galactose residue bound in the \(\alpha\)-1 subsite. This water is perfectly poised to attack the anomeric center and displace the leaving group (Fig. 4C), consistent with an inverting catalytic mechanism in which these glutamate and aspartate residues are candidate general bases. The general acid catalytic residue in these enzymes are very likely Glu\(^{558}\) in Sp3GH98 and Glu\(^{158}\) in Sp4GH98. Of both of these residues are positioned \(2.5\) Å from O1 of the reducing end galactose of the glycoside product, forming a strong hydrogen bond and allowing delivery of the proton to the glycosidic oxygen from the \(\alpha\)-hemiacetal product (Fig. 4C) (42). Furthermore, their deletion results in enzymes apparently lacking activity, as discussed above, which is consistent with assignment of these residues as the catalytic general acid residues. The enzyme-product complexes reveal, in both cases, an identical and somewhat unusual conformation of the scissile glycosidic linkage of the galactosyl-\(\alpha\)-1,4-N-acetylgalactosaminyl disaccharide fragment of the substrate that bridges the \(\alpha\)-1 subsite.
and −1 catalytic subsites. The planes of the two pyranose rings of this portion of the substrate are at roughly right angles to each other with, as discussed above, the glycosidic oxygen accepting a hydrogen bond from what we propose to be the general acid catalyst (Fig. 4C). The positioning of this catalytic group is stabilized by hydrogen bonding interactions with the tryptophan residue in the +1 subsite and Lys220 (Sp4GH98) or the structurally equivalent Lys624 (Sp3GH98). These same lysine residues also interact with the nucleophilic water molecule and may help to enhance its nucleophilicity through electrostatics or by controlling its orientation. This lysine residue probably also modulates the pKₐ values of the adjacent catalytic groups. More detailed mechanistic studies will be required to precisely identify the functional role of this residue as well as those active site residues discussed above.

Taken together, these results strongly indicate that GH98 enzymes use an inverting catalytic mechanism, which differs from previous tentative bioinformatics proposals that have suggested that these enzymes use a retaining catalytic mechanism (41).

Type 2 Carbohydrate Antigen Degradation by S. pneumoniae—Two independent studies have now shown that the fucose utilization operon is critical to the virulence of the S. pneumoniae TIGR4 strain. Four components of the operon were identified as virulence factors by signature-tagged mutagenesis: two putatively involved in carbohydrate transport, a putative fuculose kinase, and Sp4GH98 (7). In a subsequent study, deletion of the entire operon severely compromised the ability of the bacterium to cause acute respiratory disease in a mouse model (12). The biological target of the pathway encoded by this operon, however, had not been elucidated. Interestingly, Sp4GH98 is the only predicted extracellular component of this operon and thus very likely initiates this catabolic pathway by action on a host glycan. Here we have provided compelling biochemical and high resolution structural evidence that Sp4GH98 (and therefore most likely the entire pathway) is tuned to harvesting and processing the terminal H disaccharide fragment from the LewisY antigen. As we have noted, however, this particular pathway is only conserved in its entirety in 19 of the 23 available pneumococcal genomes. The other four genomes contain a variation of this fucose utilization pathway that is initiated by another family 98 glycoside hydrolase having a distinctive variation of this fucose utilization pathway that is initiated by pneumococcal genomes. The other four genomes contain a variation of this fucose utilization pathway that is initiated by pneumococcal genomes. The other four genomes contain a variation of this fucose utilization pathway that is initiated by pneumococcal genomes. The other four genomes contain a variation of this fucose utilization pathway that is initiated by pneumococcal genomes.

The biological target of the pathway encoded by this operon, however, had not been elucidated. Interestingly, Sp4GH98 is the only predicted extracellular component of this operon and thus very likely initiates this catabolic pathway by action on a host glycan. Here we have provided compelling biochemical and high resolution structural evidence that Sp4GH98 (and therefore most likely the entire pathway) is tuned to harvesting and processing the terminal H disaccharide fragment from the LewisY antigen. As we have noted, however, this particular pathway is only conserved in its entirety in 19 of the 23 available pneumococcal genomes. The other four genomes contain a variation of this fucose utilization pathway that is initiated by another family 98 glycoside hydrolase having a distinctive domain organization, here represented by Sp3GH98, which is active toward the type 2 A and B blood group glycotopes. Knowledge of the specificity of the two types of GH98 enzymes now allows us to propose a model for the harvesting and processing of fucosylated glycan antigens by S. pneumoniae (Fig. 6). In general, it is evident that the remaining components of each pathway elegantly meet the unique requirements of transporting and processing the different fucosylated glycans liberated by the two types of GH98 enzymes.

Despite the divergent substrate specificity of the GH98 enzymes initiating these two pathways, the absolute conservation in all 23 pneumococcal strains of the components dedicated to processing of glycans containing the monosaccharide fucose suggests that fucose metabolism might be a generally important feature in the pneumococcus-host relationship. The interaction of S. pneumoniae with the host involves, of course, both colonization and invasion. The observation that the operon did not contribute to the host colonization abilities in the TIGR4 strain (12) suggests that it probably features more prominently in the invasive component of the relationship of S. pneumoniae with the host. The importance of the fucose utilization pathway to virulence in other pneumococcal strains remains to be demonstrated; nevertheless, it is clear from our analysis of the GH98 enzymes from these two different pathways that pneumococci display a strain-dependent specificity for harvesting fucosylated glycans from human carbohydrate antigens bearing a type 2 core linkage.

The strain-dependent presence of the two unique types of GH98 enzymes has potential implications regarding pneumococcal virulence, assuming that these enzymes are indeed virulence factors in all or most strains of S. pneumoniae. The LewisY antigen can be found in people of all ABO blood types, whereas the A and B antigens themselves are less common (46). More specifically, the expression of the LewisY antigen in the human body is limited to certain epithelial cell types (47, 48), including the surfactant-producing type II alveolar cells of the lung (33), which are thought to be a key cell type that is targeted by S. pneumoniae (49, 50). In contrast, the A and B blood group antigens have a wide tissue distribution in hosts that have these antigens (48). Thus, the type of fucose utilization operon present in a pneumococcal strain may play a role in what cells in the host are most effectively targeted by S. pneumoniae.

More generally, this apparent relationship suggests that mismatching of the pneumococcal fucose utilization operon with the type of antigen expressed by the host (e.g. a strain with a Sp3GH98-type enzyme in a blood type O(H) individual) may impact the ability of the bacterium to establish asymptomatic colonization of the host or whether it can cause an invasive infection.

Acknowledgments—We are very thankful to Dr. Garth Ehrlich for the kind gift of the SP3-BS71 genomic DNA. We are grateful to Core D of the Consortium for Functional Glycomics for providing some of the blood group A and B tetrasaccharides. We also thank the beamline staff at the Stanford Synchrotron Research Laboratory beamline 7.1, the Canadian Light Source CMCF1, and the National Synchrotron Light Source beamline X8-C. The resources and collaborative efforts provided by the Consortium for Functional Glycomics were funded by NIGMS, National Institutes of Health, Grant GM62116.

REFERENCES
1. Tomasz, A. (1999) Ann. J. Med. 107, 555–625
2. Kadioglu, A., and Andrew, P. W. (2004) Trends Immunol. 25, 143–149
3. Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., Durkin, A. S., Gwinn, M., Kolonay, J. F., Nelson, W. C., Peterson, J. D., Umayam, L. A., White, O., Salzberg, S. L., Lewis, M. R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A. M., Utterback, T. R., Hansen, C. L., McDonald, L. A., Feldblyum, T. V., Angiuoli, S., Dickinson, T., Hickey, E. K., Holt, I. E., Loftus, B. J., Yang, F., Smith, H. O., Venter, J. C., Dougherty, B. A., Morrison, D. A., Hollingshead, S. K., and Fraser, C. M. (2001) Science 293, 498–506
4. Doern, G. V., Brueggemann, A. B., Huynh, H., and Wingert, E. (1999) Emerg. Infect. Dis. 5, 757–765
Hydrolysis of Host Carbohydrate Antigens

5. McCullers, J. A. (2006) Clin. Microbiol. Rev. 19, 571–582
6. Shelburne, S. A., Davenport, M. T., Keith, D. B., and Musser, J. M. (2008) Trends Microbiol. 16, 318–325
7. Hava, D. L., and Camilli, A. (2002) Mol. Microbiol. 45, 1389–1406
8. Chen, H., Ma, Y., Yang, J., O’Brien, C. J., Lee, S. L., Mazurkiewicz, J. E., Haataja, S., Yan, J. H., Gao, G. F., and Zhang, J. R. (2008) PLoS ONE 3, e2950
9. Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. (1998) Infect. Immun. 66, 5620–5629
10. King, S. J., Hippe, K. R., and Weiser, J. N. (2006) Trends Microbiol. 24, 221–230
11. Burnaugh, A. M., Frantz, L. J., and King, S. J. (2008) J. Bacteriol. 190, 221–230
12. Embry, A., Hinojosa, E., and Orihuela, C. J. (2007) BMC Microbiol. 7, 80
13. Obert, C., Sublett, J., Kaushal, D., Hinojosa, E., Barton, T., Tuomanen, E. I., and Orihuela, C. J. (2006) Infect. Immun. 74, 4766–4771
14. Higgins, M. A., Abbott, D. W., Boulanger, M. J., and Boraston, A. B. (2009) J. Mol. Biol. 388, 299–309
15. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2008) Nucleic Acids Res. 37, D233–D238
16. Barik, S. (1996) Methods Mol. Biol. 57, 203–215
17. Ficko-Blean, E., and Boraston, A. B. (2005) Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 61, 834–836
18. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) Nucleic Acids Res. 31, 3784–3788
19. Pflugrath, J. W. (1999) Acta Crystallogr. D Biol. Crystallogr. 55, 1718–1725
20. Powell, H. R. (1999) Acta Crystallogr. D Biol. Crystallogr. 55, 1690–1695
21. CCP4 (1994) Acta Crystallogr. D Biol. Crystallogr. 50, 760–763
22. Schneider, T. R., and Sheldrick, G. M. (2002) Acta Crystallogr. D Biol. Crystallogr. 58, 1772–1779
23. Evans, G., and Bricogne, G. (2002) Acta Crystallogr. D Biol. Crystallogr. 58, 976–991
24. Cowtan, K. D., and Zhang, K. Y. (1999) Prog. Biophys. Mol. Biol. 72, 245–270
25. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Nat. Struct. Biol. 6, 458–463
26. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132
27. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. D Biol. Crystallogr. 53, 240–255
28. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Acta Crystallogr. D Biol. Crystallogr. 61, 458–464
29. Read, R. J. (1986) Acta Crystallogr. Sect. A 42, 140–149
30. Brünger, A. T. (1992) Nature 355, 472–475
31. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) Acta Crystallogr. D Biol. Crystallogr. 55, 191–205
32. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
33. Boraston, A. B., Wang, D., and Burke, R. D. (2006) J. Biol. Chem. 281, 35263–35271
34. Gregg, K. J., Finn, R., Abbott, D. W., and Boraston, A. B. (2008) J. Biol. Chem. 283, 12604–12613
35. Anderson, K. M., Ashida, H., Maskos, K., Dell, A., Li, S. C., and Li, Y. T. (2005) J. Biol. Chem. 280, 7720–7728
36. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Biochem. J. 382, 769–781
37. Liu, Q. P., Sulzenbacher, G., Yuan, H., Bennett, E. P., Pietz, G., Saunders, K., Spence, J., Nudelman, E., Lavery, S. B., White, T., Neveu, J. M., Lane, W. S., Bourne, Y., Olsson, M. L., Henrissat, B., and Clausen, H. (2007) Nat. Biotechnol. 25, 454–464
38. Calcutt, M. J., Hsieh, H. Y., Chapman, L. F., and Smith, D. S. (2002) FEBS Microbiol Lett 214, 77–80
39. Kressin, E., and Henrick, K. (2004) Acta Crystallogr. D Biol. Crystallogr. 60, 2256–2268
40. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Biochem. J. 321, 557–559
41. Rigden, D. J. (2005) FEBS Lett. 579, 5466–5472
42. Vucadlo, D. J., and Davies, G. J. (2008) Curr. Opin. Chem. Biol. 12, 539–555
43. Shaikh, F. A., and Withers, S. G. (2008) Biochem. Cell Biol. 86, 169–177
44. Lillelund, V. H., Jensen, H. H., Liang, X., and Bols, M. (2002) Chem. Rev. 102, 515–553
45. Vasolla, A., Davies, G. J., and Böhm, M. (2002) Curr Opin Chem Biol 6, 619–629
46. Marionneau, S., Cailleau-Thomas, A., Rocher, J., Le Moullac-Vaidye, B., Ruvoën, N., Clément, M., and Le Pendu, J. (2001) Biochimie 83, 565–573
47. Yuriev, E., Farrugia, W., Scott, A. M., and Ramsland, P. A. (2005) Immunol Cell Biol. 83, 709–717
48. Rav, Y., and Dabelsteen, E. (2000) APMIS 108, 1–28
49. Cundell, D. R., and Tuomanen, E. I. (1994) Microb. Pathog. 17, 361–374
50. Lammerts van Bueren, A. L., Higgins, M., Wang, D., Burke, R. D., and Boraston, A. B. (2007) Nat. Struct. Mol. Biol. 14, 76–84