JBC Papers in Press. Published on September 24, 2018 as Manuscript RA118.003302
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.RA118.003302

Functional metagenomics identifies an exosialidase with an inverting catalytic mechanism that defines a new glycoside hydrolase family (GH156)

Léa Chuzel[1,2], Mehul B. Ganatra[1], Erdmann Rapp[2,3], Bernard Henrissat[4,5,6] and Christopher H. Taron[1]*

From [1]New England Biolabs, 240 County Road, Ipswich, MA 01938, USA; [2]Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse 1, 39106 Magdeburg, Germany; [3]glyXera GmbH, Leipziger Straße 44, 39120 Magdeburg, Germany; [4]Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, F-13288 Marseille, France; [5]INRA, USC 1408 AFMB, 13288, Marseille, France; [6]Department of Biological Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

Running title: *A new family of exosialidases with an inverting mechanism*

*To whom correspondence should be addressed: Christopher H. Taron: New England Biolabs, 240 County Road, Ipswich, MA 01938, USA; taron@neb.com; Tel.(978) 380-7207.

Keywords: exosialidase, GH156, functional metagenomics, inverting mechanism, glycoside hydrolase, glycosylation, glycan, sialic acid, neuraminidase

ABSTRACT

Exosialidases are glycoside hydrolases that remove a single terminal sialic acid residue from oligosaccharides. They are widely distributed in biology, having been found in prokaryotes, eukaryotes, and certain viruses. Most characterized prokaryotic sialidases are from organisms that are pathogenic or commensal with mammals. However, in this study, we used functional metagenomic screening to seek microbial sialidases encoded by environmental DNA isolated from an extreme ecological niche, a thermal spring. Using recombinant expression of potential exosialidase candidates and a fluorogenic sialidase substrate, we discovered an exosialidase having no homology to known sialidases. Phylogenetic analysis indicated that this protein is a member of a small family of bacterial proteins of previously unknown function. Proton NMR revealed that this enzyme functions via an inverting catalytic mechanism, a biochemical property that is distinct from those of known exosialidases. This unique inverting exosialidase defines a new CAZy glycoside hydrolase family we have designated GH156.

Enzymes are nature's biological catalysts. They perform chemical reactions with remarkable speed and specificity, making them critical for the biochemical processes that sustain life. Most enzymes have evolved to perform their function under moderate environmental conditions that support the majority of life on our planet. However, many organisms thrive in physically and chemically extreme environments such as hot springs, hydrothermal vents, hypersaline ponds, evaporation salterns, and acidic or soda lakes. These extremophiles often produce enzymes with biochemical properties that have been uniquely shaped by their extreme surroundings. As such, the study of enzymes from microbial communities that populate extreme environments is of interest to both basic science and biotechnology.

Over the past decade, functional metagenomics has become increasingly popular for identification of novel enzyme activities in nature. In a functional metagenomics workflow, high molecular weight DNA is extracted *en masse* from microbes that populate an environmental niche. This environmental DNA is cloned into a fosmid vector and introduced into a host cell, often *E. coli* (1). Individual *E. coli* clones can then be screened in high throughput for expression of a desired

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

enzyme activity using an appropriate reporter substrate. This approach is advantageous because it allows for enzyme discovery from both cultivable and uncultivable microorganisms from any environmental niche. Additionally, because the screening is based on enzyme activity and not sequence similarity to known enzymes, truly novel and unique proteins can be identified.

Functional metagenomics has been used to identify a variety of carbohydrate hydrolases (glycosidases) from environmental libraries (2–4). Several studies have reported discovery of novel glycosidases (cellulases, glucosidases, xylanases, etc.) that improve hydrolysis of plant polysaccharides with the aim of improving biomass conversion for biofuel production (3, 5–7). Additionally, novel glycosidases that improve food processing techniques have also been reported (8). Most of these studies have interrogated libraries from mesophilic microbial populations from intestinal, terrestrial or aquatic niche. Only a few studies have identified novel glycosidases from extremophiles, although cold, heat and alkaline tolerant glycosidases have been reported (9–11). The majority of glycoside hydrolase functional metagenomics studies have sought enzymes involved in degradation of plant biomass, an expected ability of many microbial populations that occupy terrestrial niche. However, the diversity of glycosidase specificities present in extremophile populations remains largely unexplored. We have an interest in enzymes that degrade complex oligosaccharides found on eukaryotic glycoproteins and glycolipids. As such, we have begun screening various extremophile libraries for activities typically associated with complex glycan degradation (*e.g.*, sialidases, galactosidases, fucosidases, mannosidases, etc). In the present study, we conducted a functional metagenomic screen of an environmental DNA library from a freshwater thermal spring (avg. temperature ~60°C) for enzymes capable of hydrolysis of sialic acids ('sialidases' or sometimes 'neuraminidases', EC 3.2.1.18).

Sialic acids are a structurally diverse family of monosaccharides that share a common nine-carbon backbone. In eukaryotes, they are generally found in a terminal position on *N*-glycans, *O*-glycans, glycosphingolipids and GPI anchor side-branches. Additionally, many bacteria that closely associate with animals incorporate sialic acids into their cell walls (12). In mammals, the most common form of sialic acid is modified at the C-5 position with an acetyl group (Neu5Ac). However, in non-human vertebrates a hydroxylated acetyl group (Neu5Gc) can occupy the C-5 position. Commonly, a sialic acid residue is attached to an adjacent neutral sugar (*e.g.*, galactose, *N*-acetylgalactosamine, *N*-acetylglucosamine) via an α2-3 or α2-6 linkage. However, other linkages are also possible, including polysialic acid chains, in which many Neu5Ac residues are each connected via an α2-8 linkage (13).

Sialidases are enzymes that catalyze hydrolysis of the glycosidic bond that links a sialic acid to a subterminal sugar in an oligosaccharide. They are widely distributed in biology and have been found in animals, fungi, protozoa, bacteria and many viruses. Based on their amino acid sequences, sialidases have been classified into four CAZy (Carbohydrate-Active Enzymes database, www.cazy.org) hydrolase families (14). Families GH33, GH34 and GH83 each contain exosialidases, enzymes that release a terminal sialic acid from oligosaccharides. In addition, family GH58 comprises endosialidases, enzymes that cleave within a polysialic acid chain (15). Families GH34 and GH83 are exclusive to viral neuraminidases, whereas family GH33 contains all known bacterial and mammalian exosialidases. Despite their lack of primary amino acid sequence similarity, all four of these sialidase families share a common 6-fold β-propeller tertiary structure (16–23). Mechanistically, known exosialidases release sialic acid with overall retention of its anomeric conformation (24–27). In contrast, endosialidases function via an inverting catalytic mechanism (28).

In the present study, we report functional metagenomic screening of a freshwater thermal hot spring fosmid library for sialidase activity. To our knowledge, this is the first functional metagenomic survey aimed at identification of sialidase activity from an extreme environment. We report the discovery and biochemical characterization of a novel exosialidase that defines a new CAZy glycoside hydrolase family (GH156). Additionally, we show that this newly found enzyme catalyzes sialic acid hydrolysis via an inverting reaction mechanism, the first example of a wild-type exosialidase with this ability.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

Results

Metagenomic DNA library construction and functional screening

A small metagenomic DNA library was constructed from environmental DNA isolated from hot spring mats collected in Dixie Valley, Nevada. The library was introduced into *E. coli* and 616 individual clones were arrayed into two 384 well plates to facilitate enzyme screening. Restriction fragment analysis of 12 randomly selected library clones indicated an average cloned insert size of ~30-40 kb (Fig. 1A). Additionally, Sanger sequencing of the ends of each insert revealed that cloned DNA originated from both known and unknown bacterial species. Among the known species were the thermophiles *Caldilinea aerophila* and *Thermomicrobium roseum* (29, 30). Other clones were related to *Chloroflexus sp*, *Thermocrinis ruber* or *Acidobacteria bacterium* and 5 clones contained DNA from unknown origins. Thus, even in a small sampling of randomly isolated clones, the library harbored genetic material from a diverse range of microbes.

To identify active sialidases, the arrayed clones were screened with 5-bromo-4-chloro-3-indolyl α-D-*N*-acetylneuraminic acid (X-Neu5Ac) and 2'-(4-methylumbelliferyl)-α-D-*N*-acetylneuraminic acid (4MU-α-Neu5Ac) substrates in agar plate and cell lysate assays, respectively. In the agar plate screen, a single colony (designated G7) hydrolyzed X-Neu5Ac and turned blue after an overnight incubation at 37°C and several days of incubation at 4°C (Fig. 1B). In the cell lysate screen, the same clone, G7, was identified by measuring an increase in fluorescence due to the hydrolysis of 4MU-α-Neu5Ac (Fig. 1C). No additional clones gave rise to detectable sialidase activity.

Sialidase gene identification

To identify the gene responsible for the observed sialidase activity, fosmid G7 was sequenced using the PacBio RS II platform. A total of 44,770 reads were generated with an average insert size of 7,523 bp and an average polymerase read length of 15,285 bp. The reads were assembled with the RS_HGAP_Assembly.3 algorithm (31). After removal of the repeat region (often observed during *de novo* assembly of circular sequences), a single contig corresponding to the full insert nucleotide sequence (41,198 bp) was obtained (GenBank accession number MH016668). The sequence coverage was approximately 8,000x. The fosmid insert did not match any DNA sequence in GenBank indicating it was from a previously unsequenced organism.

The fosmid G7 DNA sequence encoded 40 putative open reading frames (ORFs) that were predicted by MetaGeneMark (32). BLASTP analysis of the non-redundant GenBank protein sequence database was performed using each predicted ORF protein sequence as a query. Based on the observed homology of each ORF to homologous proteins in GenBank, the 40 ORFs were classified into three categories: i) proteins of known function, ii) proteins involved in saccharide utilization, and iii) hypothetical proteins of unknown function (Fig. 2A & 2B). Of the 40 ORFs, four showed similarity to proteins involved in saccharide utilization, and none showed similarity to proteins from the known CAZy sialidase families GH33, GH34, GH58 and GH83.

As sequence analysis did not identify an obvious sialidase candidate amongst the 40 sequenced ORFs, transposon mutagenesis was performed to disrupt expression of the sialidase activity. Tn5 mutagenesis was used to randomly insert a kanamycin cassette into fosmid G7 under conditions that minimized multiple insertion events. A total of 192 mutants were assayed for sialidase activity using the substrate 4MU-α-Neu5Ac (Fig. S1). Seventeen mutants with abolished sialidase activity were identified. Each of these mutants was bidirectionally Sanger-sequenced with transposon-specific primers to identify the element's insertion site. Multiple transposon insertion events were observed for 7 mutants and they were not further investigated. A map of the G7 insertion sites for the remaining 10 mutants is shown in Figure 3A. All loss-of-function insertions clustered to a ~5 kb region of fosmid G7 containing several ORFs with similarity to enzymes that hydrolyze or metabolize sugars (Fig. 2A & 3A). The majority of insertions occurred in ORF9 (annotated as encoding a putative α-L-fucosidase) or ORF12 (annotated as encoding a "protein of unknown function").

To determine if either ORF9 or ORF12 encoded a protein with sialidase activity, each was expressed using the PURExpress *in vitro* protein synthesis system (Fig. 3B) and tested for its ability to cleave 4MU-α-Neu5Ac (Fig. 3C). The expressed

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

product from ORF12 (ORF12p) generated a fluorescence signal ~27 times higher than the no DNA control reaction (the PURExpress mix alone). Expression of ORF9 yielded no sialidase activity. As the deduced protein sequence of ORF9 (ORF9p) had sequence homology to α-L-fucosidases, the ability of *in vitro* expressed ORF9p to cleave the chromogenic fucosidase substrate 2-chloro-4-nitrophenyl α-L-fucopyranoside (CNP-fucose) was also tested. Light absorbance at 405 nm was ~6 times higher for ORF9 than the negative control, confirming the ORF9 encoded an α-fucosidase, not an exosialidase. Together, these data support the conclusion that the gene product from ORF12 (Fig. 3D, GenBank accession number MH016668) is solely responsible for the hydrolysis of 4MU-α-Neu5Ac observed during primary screening. Furthermore, as ORFs 8-12 are co-directionally oriented and are each related in some way to saccharide metabolism, it is plausible that these genes are co-transcribed in the organism from which the locus originates and in *E. coli*. This could account for the observation of reduced sialidase activity upon insertion of Tn5 into multiple positions in this region.

Sialidase biochemical characterization

To facilitate biochemical characterization of ORF12p, an expression DNA construct encoding ORF12p with a C-terminal hexahistidine tag (ORF12p-His) was assembled. ORF12p-His was expressed in *E. coli* and purified using nickel affinity chromatography (Fig. 4A).

The substrate 3'-SLN-2AB was used to determine the optimum pH and temperature of the ORF12 sialidase as described in Experimental Procedures. A pH optimum of 5.0-5.5 was observed (Fig. 4B). The enzyme showed optimal activity at 37°C and retained at least 50% activity at 60°C (Fig. 4C). No metal ions were required for activity (Fig. 4D). Reaction kinetics were determined using the substrate 4MU-α-Neu5Ac. The initial reaction rates were determined at several substrate concentrations (Fig. 4E). The K_M was calculated to be 62 μM and the k_{cat} 2.3 s^{-1}.

The specificity of purified ORF12p-His was assayed using 2-aminobenzamide (2-AB)-labeled 3' and 6'-sialyl-*N*-acetyllactosamine substrates (termed 3'-SLN-2AB and 6'-SLN-2AB, respectively), a GD3 ganglioside head group glycan (Neu5Ac(α2–8)Neu5Ac(α2–3)Gal(β1–4)Glc-2AB) and G2S2 bi-antennary *N*-glycans with two terminal Neu5Gc or Neu5Ac residues (G2S2-Ac or G2S2-Gc). Digestion products were analyzed by ultra-high performance liquid chromatography with fluorescence detection (UPLC-HILIC-FLR) (Fig. 5). Negative control reactions consisting of each substrate and no enzyme yielded a single major peak for all 5 substrates (Fig. 5A, 5B, 5C, 5D & 5E top panel). A positive control reaction using commercial α2-3,6,8,9 Neuraminidase A (NeuA, NEB) resulted in a shift of each substrate, illustrating complete removal of α2-3, α2-6 and α2-8 terminal sialic acid (Fig. 5A, 5B, 5C, 5D & 5E, bottom panel). This same peak shift was observed for 3'-SNL-2AB and 6'-SNL-25AB substrates when incubated with purified ORF12p-His, indicating that this enzyme was able to hydrolyze both terminal α2-3 and α2-6 Neu5Ac (Fig. 5A & 5B, middle panel) and confirming the ability of ORF12p-His to function as an exosialidase. The peak corresponding to the released glycan from GD3 ganglioside did not shift upon incubation with ORF12p-His confirming the enzyme's inability to hydrolyze α2-8 linked sialic acid (Fig. 5C, middle panel). *N*-Glycan substrates G2S2-Ac and G2S2-Gc were both hydrolyzed by ORF12p-His indicating that this enzyme can hydrolyze both terminal Neu5Ac and Neu5Gc from complex *N*-glycan structures (Fig. 5D & 5E, middle panels). However, the enzyme showed a preference for Neu5Ac. In addition, the enzyme was assayed on a 2-AB-labelled *O*-glycan library (Fig. S2) and showed the ability to hydrolyze terminal sialic acid linked to *N*-acetylgalactosamine or galactose residues.

ORF12p sialidase reaction mechanism

Glycoside hydrolases are grouped into two mechanistic classes: inverting or retaining enzymes, depending on whether the stereochemistry of the substrate's anomeric carbon is inverted or retained upon hydrolysis. To date, known exosialidases function as retaining glycoside hydrolases, as they hydrolyze terminal Neu5Ac with net retention of its overall anomeric configuration (24–27).

Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy was used to determine if ORF12p functions via a retaining or inverting mechanism. In this experiment, purified ORF12p-His was incubated with the 4MU-α-Neu5Ac

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

substrate. Data analysis focused on the up-field region of the NMR spectra, where signals from the axial and equatorial protons in the 3-position of Neu5Ac appear. The initial NMR spectrum (Fig. 6A) showed only the axial (α-MU-H3a) and equatorial (α-MU-H3e) protons from the substrate 4MU-α-Neu5Ac at 1.99 and 2.86 ppm, respectively. The first additional set of signals (H3a and H3e) appeared after only a few minutes at 1.81 and 2.20 ppm. These signals corresponded to β-Neu5Ac that had been released from the substrate by the enzyme. The signal set of H3a and H3e corresponding to α-Neu5Ac (at 1.61 and 2.72 ppm) appeared after about 1 h and remained very small throughout the remaining reaction. This reflects spontaneous mutarotation of the β-Neu5Ac product to its α conformer and is not a result of the action of the sialidase. These data unambiguously show that the ORF12 sialidase initially produces β-NeuAc from 4MU-α-Neu5Ac and supports the conclusion that the enzyme functions via an inverting mechanism.

A control experiment was run with purified NeuA (NEB), a known retaining exosialidase from the GH33 family (33). The initial NMR spectrum (Fig. 6B) showed the axial (H3a) and equatorial (H3e) protons corresponding to 4MU-α-Neu5Ac at 2.03 and 2.89 ppm, respectively. Upon initiation of the reaction, the first additional signals appeared after about 30 minutes and corresponded to the α-anomer (1.64 and 2.76 ppm). These peaks were initially larger than the H3a and H3e β-Neu5Ac peaks (1.85 ppm and 2.03, respectively). However, after 130 min, the peaks of the β-anomer became more prominent than those of the α-anomer, consistent with the spontaneous mutarotation of the α-Neu5Ac product to its β-Neu5Ac anomer. This was previously reported to reach an equilibrium mixture of 92.1% of β-anomer and 7.5% of α-anomer and three open-chain species at pH 8.0 (34). This stereochemical pattern of product release is characteristic of retaining exosialidases and is markedly different than that of ORF12p-His, further supporting our conclusion that ORF12p-His functions via an inverting mechanism of hydrolysis.

ORF12p sialidase family phylogeny

ORF12p homologs were identified in the GenBank sequence repository using the BLASTP algorithm. The 38 nearest homologs of ORF12 were retrieved and used for study of ORF12p phylogeny. Signal peptides that were present in many sequences were trimmed along with C-terminal extensions present for two homologs (OGV64360|Lentisphaerae and KXK35835|Omnitrophica). Trimmed sequences ranging from 477 to 578 amino acids were then aligned with Muscle and a phylogenetic tree was made using BLOSSUM62 substitution matrix and the Neighbor-Joining method (Fig. 7A). It is noteworthy that all 38 ORF12p homologs showed no similarity to known CAZy glycoside hydrolases or other proteins of known function, and all were annotated as hypothetical proteins.

All sequences comprising the ORF12p family originated from bacteria and were distributed within 8 prokaryotic phyla. The majority of sequences originated from organisms in the phylum Bacteroidetes or the superphylum PVC that contains the phyla Planctomycetes, Verrumicrobia, Omnitrophica and Lentisphaerae. Other sequences were from organisms from the phyla Firmicutes, Armatimonadetes and Acidobacteria. Based on our phylogenetic analysis, ORF12p was most closely related to sequences from organisms of the PVC superphylum. Notably, no homolog was found within the phylum Proteobacteria that contains most of the well-studied bacteria, or in the phyla Actinobacteria or Cyanobacteria that contains many species widely represented in soil and aquatic environments. Similarly, no eukaryotic or archaeal proteins had similarity to ORF12p.

To determine if a second member of the ORF12p sequence family also had sialidase activity, the homologous protein OIO94155 from *Armatimonadetes* was expressed in *E. coli* and crude lysate was assayed with 4MU-α-Neu5Ac. Lysate from *E. coli* expressing ORF12p was also assayed. A high fluorescence signal was detected for both ORF12p and OIO94155 lysates indicating that OIO94155 is also an exosialidase (Fig. 7B). A multiple sequence alignment of 10 proteins (including ORF12p and OIO94155) showed the conserved regions within this family (Fig. 7A, yellow box & 7C). Strictly conserved aspartate, glutamate and histidine residues are indicated (Fig 7C, red triangles). These types of residues are commonly involved in catalysis in most inverting glycoside hydrolases (35–39). Thus, these positions will be candidates for future site-directed mutagenesis studies.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

The cloned fosmid DNA sequence encoding ORF12p was unique and showed no significant match to DNA sequences from known organisms, thus providing no clue to its precise origin. The gene encoding ORF12p resided in a gene cluster that also encoded a GH29 family α-fucosidase and a GH20 family *N*-acetylglucosaminidase (Fig. 2). While this locus did not conform to the currently defined structural features of a polysaccharide utilization locus (PUL) (40, 41), the proximity of these genes and broad specificity of ORF12p suggests that these proteins may function in a coordinated manner, perhaps in degradation of eukaryotic complex *N*-glycans that at times could be present in a hot spring (*i.e.*, decaying animal or plant matter).

Discussion

In the present study, we utilized functional metagenomic screening to search for environmental enzymes with the ability to hydrolyze the fluorogenic sialidase substrate 4MU-α-Neu5Ac. Our screen identified a 41 kb environmental DNA fragment from a hot spring metagenomic library that produced sialidase activity in *E. coli*. The DNA fragment originated from an unknown microorganism and encoded no ORFs with homology to known sialidase families. Tn5 mutagenesis was used to identify a 505 amino acid ORF responsible for the observed sialidase activity (ORF12). We undertook the phylogenetic analysis of the ORF12 protein (ORF12p) family and biochemical characterization of this new sialidase.

Glycoside hydrolases and other enzymes that assemble or modify oligo- and polysaccharides are classified into families based on their amino acid sequence similarity (14). These enzymes are catalogued by family in the curated CAZy database. Currently, there are three distinct CAZy families of exosialidases; GH33, GH34 and GH83. The ORF12p sialidase identified in this study showed no homology to any members of these exosialidase families or to members of GH58, a family of endosialidases. Searches of public sequence repositories revealed that ORF12p was a member of a small family of proteins that are distributed only amongst bacteria from ~8 phyla, several of which have been found in aquatic environments. No ORF12p family members were found in the most common bacterial phyla (*i.e.*, Proteobacteria), archaea or eukaryotes. Thus, this new sialidase family appears to be highly niche-specific in bacterial biology. The ORF12 protein sequence family defines a novel CAZy glycoside hydrolase family that has been designated GH156.

Biochemical characterization of ORF12p revealed both functional similarities and differences with known sialidases. ORF12p was capable of removing α(2,3) and α(2,6) Neu5Ac from the terminal position of oligosaccharides, indicating that it functions as an exosialidase. Despite lacking protein sequence similarity with known exosialidases, ORF12p shares similar basic biochemical properties with them (42–44). For example, ORF12p was broadly active from pH 4.5-8.0 (optimal at pH 5.0), from 20-60°C (optimal at 37°C) and did not require metal ions, all features consistent with enzymes from the GH33, GH34 and GH83 families. Additionally, its reaction kinetics (K_M = 62 μM; k_{cat} = 2.3 s^{-1}) were similar to reported values for other enzymes from the GH33 family of bacterial exosialidases (Fig. 4) (45–47).

A key difference between ORF12p and known exosialidases was its mechanism of catalysis. Glycoside hydrolases are grouped into two main mechanistic classes: inverting or retaining enzymes, depending on whether the stereochemistry of the substrate's anomeric carbon is inverted or retained upon hydrolysis. Our proton ^{1}H NMR data unambiguously showed that ORF12p liberated β-Neu5Ac as the primary product of hydrolysis of the substrate (4MU-α-Neu5Ac) (Fig. 6), indicating that it naturally functions via an inverting catalytic mechanism. This observation is unprecedented as all wild-type exosialidases that have been described to date (from prokaryotes, eukaryotes and viruses) hydrolyze terminal Neu5Ac with retention of its anomeric configuration (24–27). However, Watson *et al* showed that mutations introduced into the active site of one bacterial sialidase could induce a change in catalytic mechanism from retaining to inverting (48). The authors presented a model whereby water is able to more easily access the active site and directly acts as a nucleophile. To understand how ORF12p achieves inverting catalysis, a deeper exploration of its reaction mechanism is needed. We are currently attempting crystallization of ORF12p for x-ray structure determination to enable the characterization of the molecular determinants of substrate recognition and catalysis.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

In summary, we have identified a novel exosialidase using functional screening of a hot spring metagenomic library in *E. coli*. This enzyme is the first member of the novel CAZy family GH156 and is the first reported wild-type exosialidase to function via an inverting mechanism. This study also highlights the benefits of using function-based screening to identify unique new carbohydrate hydrolases within interesting ecological niches.

Experimental procedures

Chemicals and solvents were from Sigma-Aldrich (Saint Louis, MO). The substrate 5-bromo-4-chloro-3-indolyl-α-D-*N*-acetylneuraminic acid (X-Neu5Ac) was from Sigma-Aldrich. The substrate 2'-(4-methylumbelliferyl)-α-D-*N*-acetylneuraminic acid (4MU-α-Neu5Ac) was from Toronto Research Chemicals (North York, ON). The 2AB (2-aminobenzamide) fluorescently labeled substrates 3' and 6'-sialyl-*N*-acetyllactosamine-2AB and a 2-AB labeled di-Neu5Ac-terminated biantennary complex *N*-glycan (termed G2S2-Ac) were from Prozyme (Hayward, CA). The 2-AB labeled di-Neu5Gc-terminated biantennary complex *N*-glycan (termed G2S2-Gc) was from Tokyo Chemical Industry (Japan). All DNA enzymes and enzymes for SMRTbell complex library preparation were from New England Biolabs (NEB®; Ipswich, MA). MilliQ™ purified water (Millipore Sigma, Darmstadt, Germany) was used in all experiments.

Metagenomic DNA library construction

Samples from a hot spring located in Dixie Valley, Nevada were collected in 2013. Environmental DNA was subsequently extracted and frozen until use. Isolated DNA was concentrated to 548 ng/μL using a 0.45X volume of AMPure PB magnetic beads (Pacific Biociences, Menlo Park, CA) and used to produce a DNA library using the CopyRight v2.0 Fosmid cloning kit (Lucigen Corporation, Middleton, WI) following the manufacturer's instructions. Briefly, DNA was end repaired and size selected using a 1% low melting point agarose gel run overnight at 35 V. DNA fragments from 30-70 kb were extracted from the gel using 1 U of β-agarase I (NEB) for each 100 μL of melted agarose. The end-repaired and size-selected DNA was ligated to the pSMART FOS cloning vector. Phage packaging of the resulting clones was performed using Gigapack III XL packaging extracts (Agilent Technologies, Santa Clara, CA). Replicator FOS cells were transfected with the packaging reaction and plated on YT-CXIS agar medium (8 g Bacto-tryptone, 5 g yeast extract, 5 g NaCl, 15 g agar per liter containing 12.5 μg/mL of chloramphenicol, 40 μg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and 5% (w/v) sucrose) and incubated overnight at 37°C. A total of 616 colonies were harvested and archived in two 384 well plates in 20% sterile glycerol (v/v).

DNA library quality assessment

Restriction digestion was used to assess the average insert size in the fosmid DNA library. Fosmids were individually isolated from 12 randomly selected library clones using the FosmidMAX™ DNA purification kit (Lucigen Corporation). Each fosmid was digested with SbfI (NEB), a unique restriction site in the pSMART FOS vector. Digestion products were separated by electrophoresis on a 1% agarose gel to estimate the average insert size of cloned DNA.

To assess the cloned DNA insert species diversity, the 12 randomly selected clones were subjected to Sanger sequencing using the T7 universal primer or pCC1 forward and reverse sequencing primers (Epicentre, Madison, WI) (Table S2). These primers allowed sequencing of ~500-800 bp from the insert ends. BLASTN was used to probe the GenBank sequence repository for known identical sequences to assist in identification of the species of origin for the cloned DNA fragments.

Screening for sialidase activity

The metagenomic library was screened for sialidase activity using two different approaches. In an agar plate assay, library clones were spotted onto agar plates (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g dextrose, 1 g $MgCl_2$-$6H_2O$, 20 g bacto-agar, and 2 mL 2M NaOH per liter) supplemented with 60 μg/mL X-Neu5Ac dissolved in dimethylformamide (DMF) using a 384 Slot Pin Multi-Blot™ Replicator (V&P Scientific Incorporation, San Diego, CA). After overnight incubation at 37°C, the plates were stored at 4°C until blue color developed (several days). For sialidase activity screening in *E. coli* cell lysates,

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

the library clones were grown in 50 µL LB liquid cultures (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g dextrose, 1 g $MgCl_2$-$6H_2O$, 2 mL of 2M NaOH per liter, containing chloramphenicol 12.5 µg/mL and 1X inducing solution (Lucigen Corporation)) in 384 well plates and incubated overnight at 37°C. Fifty microliters Y-per lysis buffer (Thermo Fischer Scientific, Waltham, MA) containing 40 µg/mL of the sialidase substrate 4MU-α-Neu5Ac were added to the liquid cultures. The mixtures were incubated overnight at 37°C in a static incubator. Fluorescence at λ_{ex} = 365 nm and λ_{em} = 445 nm was read with a SpectraMax Plus 384 Microplate Reader (Molecular Devices, Sunnyvale, CA).

Pacific Biosciences sequencing of fosmid clones

Fosmid DNA was extracted using the FosmidMAX™ DNA purification kit (Lucigen Corporation) following the manufacturer's instructions. DNA was isolated from 40 mL LB cultures supplemented with 12.5 µg/mL chloramphenicol and 1X inducing solution (Lucigen Corporation). A 10-kb SMRTbell library was prepared as follows. Ten micrograms of DNA in 150 µL TE buffer (10 mM Tris pH 8.0, 0.1 mM EDTA) were sheared using g-tubes (Covaris, Woburn, MA). The g-tubes were centrifuged at 4500 r.p.m. (Eppendorf 5424 centrifuge) for 30 seconds cap up, followed by 30 seconds cap down, three times each for a total of 6 spins. Sheared DNA was concentrated using 0.45X volume of AMPure PB magnetic beads (Pacific Biosciences). The beads were incubated for 10 min in a ThermoMixer C (Eppendorf, Hamburg, Germany) at 1,400 r.p.m. then washed twice on a magnetic rack with 500 µL of freshly prepared 70% ethanol. DNA was eluted from the beads by addition of 20 µL of TE buffer and incubation for 2 min in a ThermoMixer C at 1,400 r.p.m. Two microliters of PreCR Repair Mix (NEB) and 5 µL of NEBNext End Repair Reaction Buffer (NEB) in a 50 µL total volume reaction was added to repair the damaged DNA template. The reaction was incubated at 37°C for 20 min. DNA was end-repaired by addition of 2.5 µL NEBNext End repair mix (NEB) (incubation at 25°C for 5 min) and then purified with 0.45X volume of AMPure PB magnetic beads and eluted with 20 µL of TE. DNA was blunt-ligated with 2.5 µL of 80 µM SMRT bell adaptors (Integrated DNA Technologies, San Jose, CA) using 17.5 µL of Blunt/TA Ligase Master Mix (NEB). The ligation reaction was incubated at 25°C for 15 min. Ligase was removed using the Monarch PCR and DNA Cleanup kit (NEB). Unligated DNA fragments were removed by digestion with Exonuclease III and VII (100 U each) in the presence of 4 µL of buffer for T4 DNA ligase with 10 mM ATP. Finally, two AMPure PB magnetic bead purifications were performed. The SMRTbell library was ultimately eluted with 15 µL TE buffer. MagBead complexes were prepared on the day of sequencing using the Binding Calculator tool and enzymes from Pacific Biosciences. The final library was sequenced on a PacBio RS II platform using the P6 chemistry and one single-molecule real-time (SMRT) cell for a 360-min movie.

Reads were assembled *de novo* using the RS_HGAP_Assembly.3 protocol (PacBio Portal). ORFs encoded by the assembled contig were predicted using MetaGeneMark (32). The deduced protein sequence of each ORF was subjected to BLASTP analysis against the GenBank non-redundant protein database. ORFs that showed good identity and coverage scores to several proteins of the same function were annotated with this putative function. ORFs that showed no significant homology to any protein of known function or that matched proteins annotated as "hypothetical proteins" were assigned as such for our study.

Tn5 mutagenesis

A library of random mutants of fosmid G7 was generated using the EZ-Tn5™ <Kan-2> Insertion kit (Lucigen Corporation) following the manufacturer's instructions. Briefly, an equimolar amount (0.007 pmol) of the EZ-Tn5™ <KAN-2> transposon and target DNA (G7 fosmid) were used to maximize the insertion efficiency while minimizing multiple insertion events. The 10 µL reaction was incubated at 37°C for 2 h and stopped upon addition of 1 µL EZ-Tn5 10X stop solution and incubation at 70°C for 10 min. One microliter of the reaction was added to 50 µL of thawed electrocompetent cells EC300110 (Lucigen Corporation) in a pre-chilled tube. Electroporation was performed with a Gene Pulser Xcell™ Electroporation System (Bio-Rad, Hercules, CA) following the manufacturer's instruction (1 mm gap, 1800 V, 25 µF and 200 Ω). SOC medium (NEB) was added immediately after electroporation (950 µL) and the cells were transferred to a 15 mL

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

tube for incubation at 37°C with shaking for 1h. Transformed cells were plated on LB + 12.5 µg/mL chloramphenicol and 50 µg/mL kanamycin. One hundred ninety-two random mutants were arrayed in two 96 well plates. Sialidase activity in both plates was investigated using the lysis-based assay with 4MU-α-Neu5Ac substrate as described above.

In vitro and in vivo sialidase expression

Sialidase candidates ORF9 and ORF12 were each expressed *in vitro* using the PURExpress *In Vitro* Protein Synthesis Kit (NEB) following the manufacturer's instruction. PURExpress DNA templates were generated by PCR using primers specific to ORF9 or ORF12 (Table S2). PCR was performed using 250 ng of fosmid G7 as template, 0.75 µL of each primer (20 µM), 15 µL Q5 Hot Start High-Fidelity 2X Master Mix (NEB) in a 30 µL total reaction volume. Thermocycling consisted of 25 cycles (98°C for 10 sec, 72°C for 30 sec, 72°C for 90 sec). The amplified product was purified using the Monarch PCR Clean-up Kit (NEB). For *in vitro* protein synthesis, 1 µL of amplified ORF9 or ORF12 DNA was mixed with 10 µL solution A, 7.5 µL solution B and 0.5 µL RNase Inhibitor Murine (NEB) and incubated for 2 h at 37°C to express the desired protein. Expression was verified by separating 2.5 µL of the reaction on a Novex 10-20% Tris-Glycine gel (Thermo Fisher Scientific, Waltham, MA). Additionally, sialidase activity of *in vitro* expressed proteins was assessed by incubating 20 µL of PURExpress product with 4 µL of 100 µg/mL 4MU-α-Neu5Ac at 37°C for 1h and reading fluorescence at λ_{ex} = 365 nm and λ_{em} = 445 nm in a SpectraMax microplate fluorometer.

ORF12 was fused with a hexahistidine tag coding sequence at its 3' end (for a C-terminal tag) and cloned into the pJS119K vector (49). Primers were designed using the NEBuilder assembly online tool (NEB) (Table S2). Linearization of the vector was performed by PCR with the Q5 Hot Start High-Fidelity 2X Master Mix (NEB) for 25 cycles (98°C for 10 sec, 60°C for 30 sec, 72°C for 2 min). An insert was also prepared by PCR using 25 cycles of 98°C for 10 sec, 60.5°C for 30 sec, 72°C for 2 min with the Q5 Hot Start High-Fidelity 2X Master Mix. The construct was assembled using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB) and transfected into NEB 5-alpha competent *E. coli* cells following the manufacturer's instructions. The clone was verified by Sanger sequencing and used to transform NEB Express competent cells. Expression under the Ptac promoter was performed at 18°C overnight upon addition of IPTG to 0.4 mM. Cells from 1L of LB culture were harvested by centrifugation at 15,000 x g for 10 min at 4°C. Cell paste was resuspended in 30 mL of 20 mM sodium phosphate, pH 7.4, 500 mM NaCl and 20 mM imidazole buffer and lysed using a TS Benchtop Series cell disruptor (Constant Systems Limited, Daventry, UK) at 32 kPsi. Expressed protein was purified on a 5 mL His-Trap™ FF column (GE Healthcare, Little Chalfont, UK). The bound ORF12-6xHis protein was eluted with 20 mM sodium phosphate, pH 7.4, 500 mM NaCl and 500 mM imidazole by gradient elution over 20 column volumes (from 0 to 100 % of elution buffer). Fractions containing pure protein were pooled and dialyzed against 20 mM sodium phosphate, pH 7.4, containing 500 mM NaCl, 1 mM EDTA.

Sialidase biochemical characterization

One unit of the ORF12 sialidase was defined as the amount of enzyme required to cleave Neu5Ac from 1 nmol of 4MU-α-Neu5Ac in 1 hour at 37°C in 20 mM sodium phosphate pH 7.4.

The effect of pH and temperature on enzyme activity was investigated by incubating 0.5 U of enzyme with 1.7 pmol of 3'-sialyl-*N*-acetyllactosamine-2AB for 3h at 37°C. These conditions led to incomplete cleavage so that both substrate and product could be monitored. A range of pH from 4.5 to 9.5 (sodium acetate 50 mM buffer for pH between 4.5 and 5.5, sodium phosphate 20 mM buffer for pH between 5.5 and 8.0 and Tris-HCl 50 mM buffer for pH between 8.0 and 9.5) at 37°C and temperature ranges from 15°C to 70°C in sodium acetate 50 mM, pH 5.0 were tested. The effect of metal ions was tested by incubating 0.5 U of enzyme with 1.7 pmol of 3'-sialyl-*N*-acetyllactosamine-2AB in 50 mM sodium acetate, pH 5.0 with 5 mM of $NiSO_4$, $CaCl_2$, $MnSO_4$, $MgCl_2$, $FeSO_4$, $ZnCl_2$ or $CuSO_4$, respectively, for 3h at 37°C. After incubation, reactions were dried by vacuum evaporation and resuspended in 1.8 µL water and 13.2 µL acetonitrile for a 12:88 ratio. Twelve microliters were injected into a Waters Acquity BEH glycan amide column (2.1 X 150 mm, 1.7 µm) on a Waters ACQUITY UPLC H-Class instrument (Waters Corporation, Milford, MA) equipped with a quaternary solvent manager and a

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

fluorescence detector. Fifty millimolar ammonium formate buffer pH 4.4 and 100 % acetonitrile were used, respectively, as solvent A and B. The gradient used was 0-1.50 min, 12% solvent A; 1.5-35 min, 47% solvent A; 35-36 min, 70% solvent A; 36.5-42 min, 12% solvent A with a flow rate of 0.561 mL/min. Samples were kept at 5°C prior to injection and separation was performed at 30°C. The fluorescence detection wavelengths were λex = 330 nm and λem = 420 nm with a data collection rate of 20 Hz. The quantity of uncleaved substrate and released product were analyzed with Empower 3 chromatography workstation software (Waters Corporation). Peak areas were quantitated by integration and relative activity was calculated.

Sialidase specificity was assessed by testing its efficacy on 2-AB-labeled 3' and 6'-sialyl-*N*-acetyllactosamine substrates (termed 3'-SLN-2AB and 6'-SLN-2AB, respectively), G2S2-Ac, G2S2-Gc and the GD3 ganglioside released glycan (Neu5Ac(α2–8)Neu5Ac(α2–3)Gal(β1–4)Glc-2AB). The released glycan head group substrate from GD3 ganglioside was prepared as follows: glycans from 10 nmol of GD3 ganglioside dissolved in methanol were released with 20 mU of endoglycoceramidase I (EGCase I, NEB) in a 10 µL reaction and incubated for 24h at 37°C. The reaction mix was passed through a Nanosep 10K Omega centrifugal device (Pall Corporation, Westborough, MA) and centrifuged at 12,000 r.p.m. for 4 min to remove the enzyme. The filtrate was dried in a vacuum evaporator prior to 2-AB labeling for 2 h at 65°C with 10 µL of 2-AB labeling mix (350 mM 2-AB, 1 M sodium cyanoborohydride in 7:3 DMSO:acetic acid). Labeled glycans were cleaned-up with a HILIC detergent removal Microspin cartridge (The Nest Group Inc., Southborough, MA).

Enzyme (1 or 10 U) was mixed with each of the substrates in 20 µL reactions in 50 mM sodium acetate pH 5.0 and incubated at 37°C overnight. Negative and positive control experiments were performed in the same conditions with no enzyme and 1 µL of α2-3,6,8,9 Neuraminidase A (NeuA, NEB) respectively. Reactions were dried in a vacuum evaporator and analyzed by UPLC-HILIC-FLR as described above. ORF12p-His activity was also tested on a fetuin *O*-glycan library purchased from Ludger (Oxfordshire, UK) and 2-AB labelled as described above.

Reaction kinetics

To determine the kinetic parameters of ORF12p-His, 20 U of ORF12p-His was incubated at 37°C with different concentrations of 4MU-α-Neu5Ac (60 - 280 µM) in 50 mM sodium acetate, pH 5.0. Each substrate concentration was tested in triplicate. Aliquots (50 µL) of each reaction were removed at different time points over a 5 min incubation. They were immediately mixed with 50 µL of 1M sodium carbonate (pH 10.9) to stop the reaction. Fluorescence was read at λ_{ex} = 365 nm and λ_{em} = 445 nm in a SpectraMax microplate fluorometer.

NMR spectroscopy

NMR was performed by the Complex Carbohydrate Research Centre (CCRC, Georgia).A solution was prepared in a 5-mm NMR tube with 320 µL 1.78 mg/mL 4MU-α-NeuAc in D_2O (2 mM final concentration after addition of enzyme), 60 µL of 200 mM sodium phosphate in D_2O, pH 7.4 (20 mM final concentration after addition of enzyme), and 190 µL D_2O. The solution was mixed and the NMR tube placed into a 600 MHz Inova NMR spectrometer (Agilent Technologies, Santa Clara, CA) at 25°C. A 1D proton NMR experiment with water presaturation (using the 2-step purge option) was recorded as the t=0 measurement. Then, 30 U ORF12p-His sialidase or 240 U of NeuA (NEB) was added, the NMR tube was inverted several times to ensure good mixing, and the tube was replaced into the NMR spectrometer. 1D proton spectra with water presaturation were recorded at time intervals with 32 transients each. Chemical shifts were referenced relative to the residual HDO peak, set at 4.78 ppm.

Armatimonadetes homolog expression

The gene encoding the hypothetical protein OIO94155 from *Armatimonadetes* was synthesized by Genscript (Piscataway, NJ) in pUC57. The gene was sub-cloned into pJS119K using the HiFi DNA Assembly Cloning Kit (NEB) with primers designed with NEBuilder assembly online tool (Table S1). The clone was verified by Sanger sequencing and used to transform NEB Express competent cells. Expression under the Ptac promoter was performed at 30°C for 4 h upon addition of IPTG to 0.4 mM. Crude lysate was

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

prepared by sonication and assayed with 4-MU-α-Neu5Ac by mixing 5 µL of crude lysate with 4 µL of 100 µg/mL 4MU-α-Neu5Ac and 15 µL of 20 mM MES buffer pH 6.5. After 1h incubation at 37°C the fluorescence of the reaction was read at λ_{ex} = 365 nm and λ_{em} = 445 nm in a SpectraMax microplate fluorometer. Lysate from NEB Express cells containing the empty pJS119K vector and lysate from the pJS119K-ORF12p construct were prepared and assayed in the same conditions to serve as negative and positive controls, respectively.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

Acknowledgements: The authors thank Dr. Donald Comb for basic research funding at New England Biolabs. We thank Rick Morgan for kindly providing hot spring environmental DNA. We also thank Barbara Taron, Saulius Vainauskas, Bill Jack and Tom Evans for comments and critical review of the manuscript prior to its submission. NMR analysis was performed by the Complex Carbohydrate Research Center supported by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, U.S. Department of Energy grant (DE-SC0015662) to Parastoo Azadi.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

References

1. Leis, B., Angelov, A., and Liebl, W. (2013) Screening and Expression of Genes from Metagenomes. *Adv. Appl. Microbiol.* **83**, 1–68
2. Cheng, J., Romantsov, T., Engel, K., Doxey, A. C., Rose, D. R., Neufeld, J. D., and Charles, T. C. (2017) Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences. *PLoS One*. **12**, 1–20
3. Mewis, K., Armstrong, Z., Song, Y. C., Baldwin, S. A., Withers, S. G., and Hallam, S. J. (2013) Biomining active cellulases from a mining bioremediation system. *J. Biotechnol.* **167**, 462–471
4. Coughlan, L. M., Cotter, P. D., Hill, C., and Alvarez-Ordóñez, A. (2015) Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Front. Microbiol.* **6**, 1–22
5. Maruthamuthu, M., Jiménez, D. J., Stevens, P., and van Elsas, J. D. (2016) A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveilsnovel thermoalkaliphilic enzymes. *BMC Genomics*. **17**, 86
6. Gladden, J. M., Eichorst, S. A., Hazen, T. C., Simmons, B. A., and Singer, S. W. (2012) Substrate perturbation alters the glycoside hydrolase activities and community composition of switchgrass-adapted bacterial consortia. *Biotechnol. Bioeng.* **109**, 1140–1145
7. Rooks, D. J., McDonald, J. E., and McCarthy, A. J. (2012) Metagenomic Approaches to the Discovery of Cellulases. in *Methods in enzymology*, pp. 375–394, **510**, 375–394
8. Sathya, T. A., and Khan, M. (2014) Diversity of glycosyl hydrolase enzymes from metagenome and their application in food industry. *J. Food Sci.* **79**, R2149–R2156
9. Mirete, S., Nica Morgante, V., and Gonzá Lez-Pastor, J. E. (2016) Functional metagenomics of extreme environments. 10.1016/j.copbio.2016.01.017
10. Vester, J. K., Glaring, M. A., and Stougaard, P. (2014) Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. *Microb. Cell Fact.* 10.1186/1475-2859-13-72
11. DeCastro, M.-E., Rodríguez-Belmonte, E., and González-Siso, M.-I. (2016) Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes. *Front. Microbiol.* **7**, 1521
12. Vimr, E. R., Kalivoda, K. A., Deszo, E. L., and Steenbergen, S. M. (2004) Diversity of microbial sialic acid metabolism. *Microbiol. Mol. Biol. Rev.* **68**, 132–53
13. Varki, A., Cummings, R. D., Esko, J., Freeze, H., Stanley, P., Bertozzi, C., Hart, G., and Etzler, M. (1999) *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, 10.1016/S0962-8924(00)01855-9
14. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495
15. Petter, J. G., and Vimr, E. R. (1993) Complete nucleotide sequence of the bacteriophage K1F tail gene encoding endo-N-acylneuraminidase (endo-N) and comparison to an endo-N homolog in bacteriophage PK1E. *J. Bacteriol.* **175**, 4354–4363
16. Varghese, J. N., Laver, W. G., and Colman, P. M. (1983) Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature*. **303**, 35–40
17. Burmeister, W. P., Henrissat, B., Bosso, C., Cusack, S., and Ruigrok, R. W. H. (1993) Influenza B virus neuraminidase can synthesize its own inhibitor. *Structure*. **1**, 19–26
18. Crennell, S. J., Garman, E. F., Laver, W. G., Vimr, E. R., and Taylor, G. L. (1993) Crystal structure of a bacterial sialidase (from Salmonella typhimurium LT2) shows the same fold as an influenza virus neuraminidase. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 9852–9856
19. Crennell, S., Garman, E., Laver, G., Vimr, E., and Taylor, G. (1994) Crystal structure of Vibrio cholerae neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure*. **2**, 535–544
20. Crennell, S., Takimoto, T., Portner, A., and Taylor, G. (2000) Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. *Nat. Struct. Biol.* **7**, 1068–1074
21. Luo, Y., Li, S. C., Chou, M. Y., Li, Y. T., and Luo, M. (1998) The crystal structure of an

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

intramolecular trans-sialidase with a NeuAc alpha2-->3Gal specificity. *Structure*. **6**, 521–30
22. Buschiazzo, A. (2000) Structural basis of sialyltransferase activity in trypanosomal sialidases. *EMBO J.* **19**, 16–24
23. Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C., and Alzari, P. M. (2002) The crystal structure and mode of action of trans-sialidase, a key enzyme in Trypanosoma cruzi pathogenesis. *Mol. Cell*. **10**, 757–768
24. Wilson, J. C., Angus, D. I., and Von Itzstein, M. (1995) 1H NMR Evidence That Salmonella typhimurium Sialidase Hydrolyzes Sialosides with Overall Retention of Configuration. *J. Am. Chem. Soc.* **117**, 4214–4217
25. Chong, A. K. J., Pegg, M. S., Taylor, N. R., and von Itzstein, M. (1992) Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. *Eur. J. Biochem.* **207**, 335–343
26. Friebolin, H., Baumann, W., Keilich, G., Ziegler, D., Brossmer, R., and von Nicolai, H. (1981) [1H-NMR spectroscopy--a potent method for the determination of substrate specificity of sialidases (author's transl)]. *Hoppe Seylers Z Physiol Chem*. **362**, 1455–1463
27. Kao, Y. H., Lerner, L., and Warner, T. G. (1997) Stereoselectivity of the Chinese hamster ovary cell sialidase: Sialoside hydrolysis with overall retention of configuration. *Glycobiology*. **7**, 559–563
28. Morley, T. J., Willis, L. M., Whitfield, C., Wakarchuk, W. W., and Withers, S. G. (2009) A new sialidase mechanism: Bacteriophage K1F endosialidase is an inverting glycosidase. *J. Biol. Chem.* **284**, 17404–17410
29. Sekiguchi, Y., Yamada, T., Hanada, S., Ohashi, A., Harada, H., and Kamagata, Y. (2003) Anaerolinea thermophila gen. nov., sp. nov. and Caldilinea aerophila gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain bacteria at the subphylum level. *Int. J. Syst. Evol. Microbiol.* **53**, 1843–1851
30. Wu, D., Raymond, J., Wu, M., Chatterji, S., Ren, Q., Graham, J. E., Bryant, D. A., Robb, F., Colman, A., Tallon, L. J., Badger, J. H., Madupu, R., Ward, N. L., and Eisen, J. A. (2009) Complete genome sequence of the aerobic CO-oxidizing thermophile Thermomicrobium roseum. *PLoS One*. **4**, e4207
31. Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., and Korlach, J. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods*. **10**, 563–569
32. Zhu, W., Lomsadze, A., and Borodovsky, M. (2010) Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, e132–e132
33. Friebolin, H., Brossmer, R., Keilich, G., Ziegler, D., and Supp, M. (1980) [1H-NMR-spectroscopic evidence for the release of N-acetyl-alpha-D-neuraminic acid as the first product of neuraminidase action (author's transl)]. *Hoppe. Seylers. Z. Physiol. Chem.* **361**, 697–702
34. Chan, J., Sandhu, G., and Bennet, A. J. (2011) A mechanistic study of sialic acid mutarotation: Implications for mutarotase enzymes. *Org. Biomol. Chem.* **9**, 4818
35. Ghate, A. A., and Air, G. M. (1998) Site-directed mutagenesis of catalytic residues of influenza virus neuraminidase as an aid to drug design. *Eur. J. Biochem.* **258**, 320–31
36. McCarter, J. D., and Stephen Withers, G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885–892
37. Vuong, T. V., and Wilson, D. B. (2010) Glycoside hydrolases: Catalytic base/nucleophile diversity. *Biotechnol. Bioeng.* **107**, 195–205
38. Czjzek, M., and Michel, G. Innovating glycoside hydrolase activity on a same structural scaffold. 10.1073/pnas.1704802114
39. Davies, G. J., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*. **3**, 853–859
40. Martens, E. C., Koropatkin, N. M., Smith, T. J., and Gordon, J. I. (2009) Complex glycan

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284**, 24673–7
41. Foley, M. H., Cockburn, D. W., and Koropatkin, N. M. (2016) The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cell. Mol. Life Sci.* **73**, 2603–17
42. Wang, X., Long, H., Shen, D., and Liu, L. (2017) Cloning, expression, and characterization of a novel sialidase from Brevibacterium casei. *Biotechnol. Appl. Biochem.* **64**, 195–200
43. Guo, J., Wang, Y., Song, B., Wang, X., Yang, G., and Guan, F. (2015) Identification and functional characterization of intracellular sialidase NeuA3 from Streptomyces avermitilis. *Process Biochem.* **50**, 752–758
44. Park, K. H., Kim, M. G., Ahn, H. J., Lee, D. H., Kim, J. H., Kim, Y. W., and Woo, E. J. (2013) Structural and biochemical characterization of the broad substrate specificity of Bacteroides thetaiotaomicron commensal sialidase. *Biochim. Biophys. Acta - Proteins Proteomics*. **1834**, 1510–1519
45. Minami, A., Ishibashi, S., Ikeda, K., Ishitsubo, E., Hori, T., Tokiwa, H., Taguchi, R., Ieno, D., Otsubo, T., Matsuda, Y., Sai, S., Inada, M., and Suzuki, T. (2013) Catalytic preference of Salmonella typhimurium LT2 sialidase for N-acetylneuraminic acid residues over N-glycolylneuraminic acid residues. *FEBS Open Bio*. **3**, 231–236
46. Useh, N. M., Ajanusi, J. O., Esievo, K. A. N., and Nok, A. J. (2006) Characterization of a sialidase (neuraminidase) isolated from Clostridium chauvoei (Jakari strain). *Cell Biochem. Funct.* **24**, 347–352
47. Jers, C., Guo, Y., Kepp, K. P., and Mikkelsen, J. D. (2015) Mutants of micromonospora viridifaciens sialidase have highly variable activities on natural and non-natural substrates. *Protein Eng. Des. Sel.* **28**, 37–44
48. Watson, J. N., Dookhun, V., Borgford, T. J., and Bennet, A. J. (2003) Mutagenesis of the Conserved Active-Site Tyrosine Changes a Retaining Sialidase into an Inverting Sialidase. *Biochemistry*. **42**, 12682–12690
49. Fürste, J. P., Pansegrau, W., Frank, R., Blöcker, H., Scholz, P., Bagdasarian, M., and Lanka, E. (1986) Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. *Gene*. **48**, 119–31
50. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F., Lütteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K. F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015) Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*. **25**, 1323–4

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

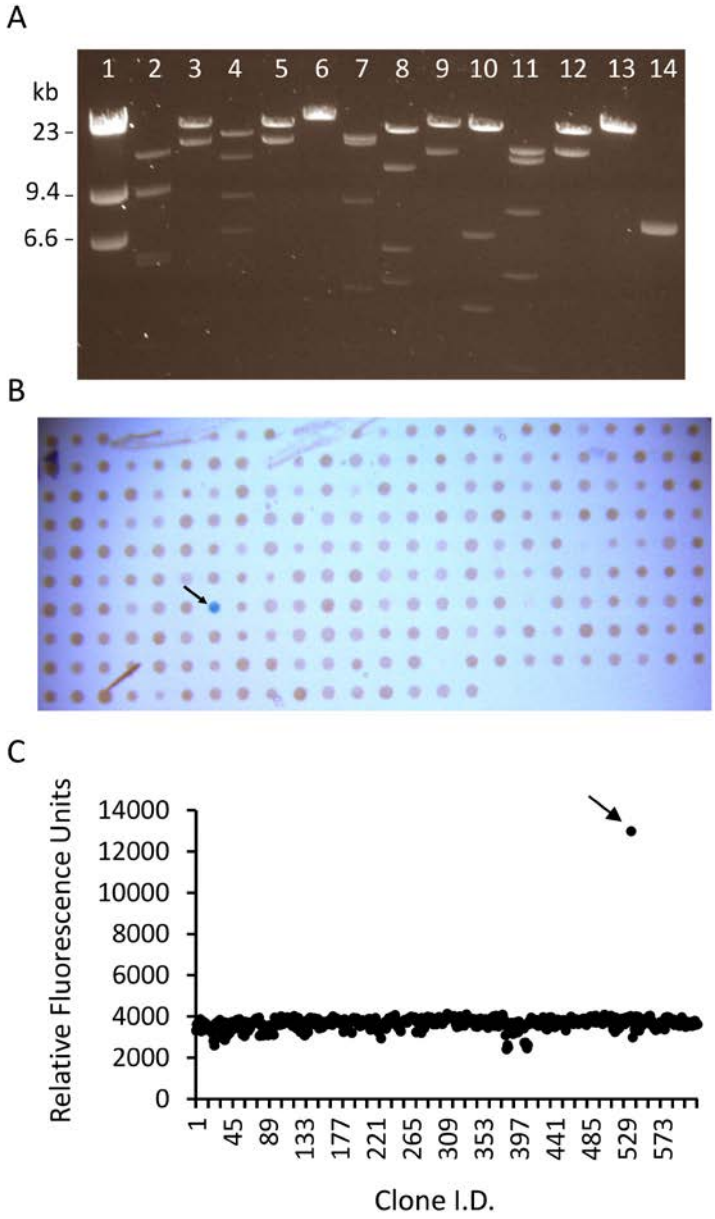
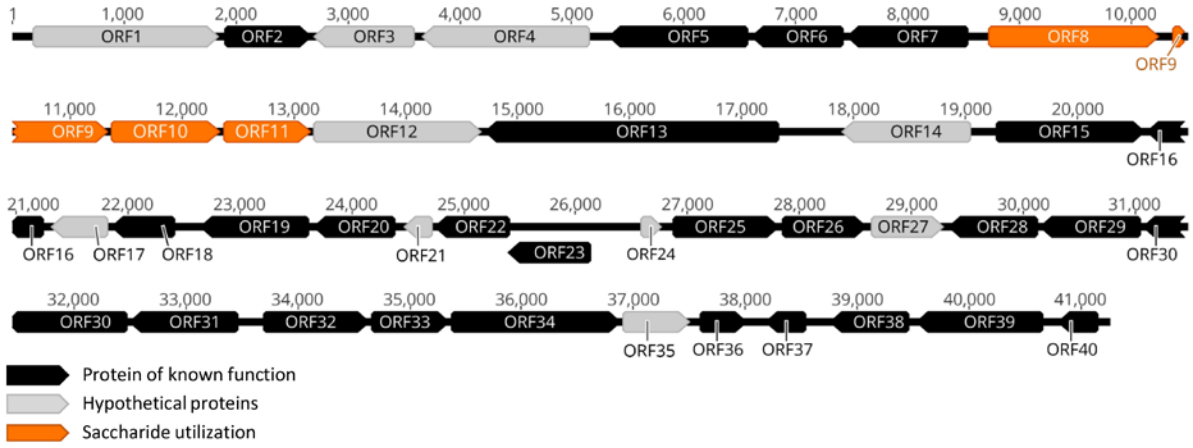


Figure 1. Screening for sialidase activity from a hot spring metagenomic library. (A) Restriction fragment analysis of 12 randomly selected clones from the hot spring metagenomic library were isolated and digested with the rare-cutting endonuclease SbfI. Digested fosmids were separated overnight on a 1% agarose gel along with a λHind III size marker and a linearized pSMART FOS empty vector control (lane 14; contains one SbfI site). Each clone showed a unique banding pattern, with fragments whose combined sizes indicated the presence of an insert of at least 30-40 kb. (B) *E. coli* cells harboring individual fosmid clones were assayed for sialidase activity with X-Neu5Ac incorporated into agar medium. A single positive clone forming a blue colony is denoted with an arrow. (C) Lysate from micro-cultures of *E. coli* cells harboring individual fosmid clones were assayed for sialidase activity with 4MU-α-Neu5Ac. A single positive clone is denoted with an arrow.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

A



B

ORF1	Hypothetical protein	**ORF11**	Xylose isomerase	**ORF21**	Hypothetical protein – possible transcriptional regulator	**ORF31**	Aspartate carbamoyltransferase
ORF2	Short-chain dehydrogenase	**ORF12**	Hypothetical protein	**ORF22**	Methyltransferase	**ORF32**	ABC transporter
ORF3	Hypothetical protein	**ORF13**	ATPase	**ORF23**	Thiamine pyrophosphokinase	**ORF33**	ABC transporter permease
ORF4	Hypothetical protein	**ORF14**	Hypothetical protein	**ORF24**	Hypothetical protein	**ORF34**	ABC transporter
ORF5	Radical SAM protein	**ORF15**	Metallo hydrolase	**ORF25**	Acetamidase	**ORF35**	Hypothetical protein
ORF6	Cytochrome c-type biogenesis protein	**ORF16**	Phosphoesterase	**ORF26**	Integrase	**ORF36**	Ureidoglycolate hydrolase
ORF7	Cytochrome c-type biogenesis protein	**ORF17**	Hypothetical protein – possible RNA binding protein	**ORF27**	Hypothetical protein	**ORF37**	Hypothetical protein
ORF8	Glycoside hydrolase family 20 (N-acetylglucosaminidase)	**ORF18**	ABC transporter	**ORF28**	Dihydroorotate dehydrogenase	**ORF38**	Creatininase
ORF9	Glycoside hydrolase family 29 (Alpha-fucosidase)	**ORF19**	ABC transporter	**ORF29**	Dihydroorotate dehydrogenase	**ORF39**	Dihydroorotase
ORF10	Glycerate dehydrogenase	**ORF20**	ABC transporter permease	**ORF30**	Orotidine decarboxylase	**ORF40**	Enamine deaminase

Figure 2. Predicted ORFs encoded in the fosmid G7 nucleotide sequence. (A) Fosmid G7 was sequenced on the PacBio RSII platform. ORFs were predicted with MetaGeneMark and classified into three categories based on their homology to proteins from the non-redundant protein database (NCBI). *Black*, proteins with a known annotated function; *Gray*, "hypothetical proteins" with no annotated function; *Orange*, proteins involved in saccharide utilization. (B) Database annotations for all 40 ORFs.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

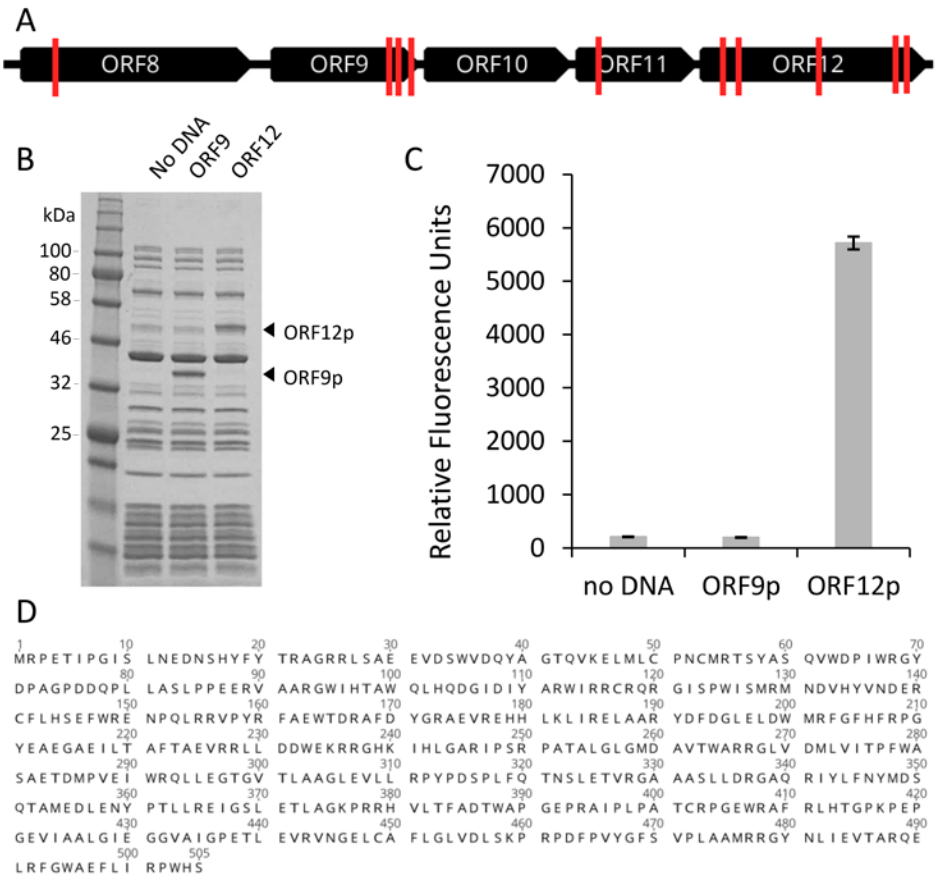


Figure 3. Identification of the sialidase-encoding ORF on fosmid G7 and its *in vitro* expression. (A) A map of fosmid G7 transposon insertion sites (red lines) in mutants with abolished sialidase activity. (B) SDS-PAGE of ORF9 and ORF12 proteins expressed *in vitro* using the PURExpress system. (C) Sialidase activity produced in PURExpress reaction mixtures was assessed using the substrate 4MU-α-Neu5Ac as described in Experimental Procedures. (D) The deduced amino acid sequence of ORF12p. Nucleotide sequence and the deduced protein sequence for ORF12 are annotated in the fosmid G7 sequence record (GenBank MH016668).

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

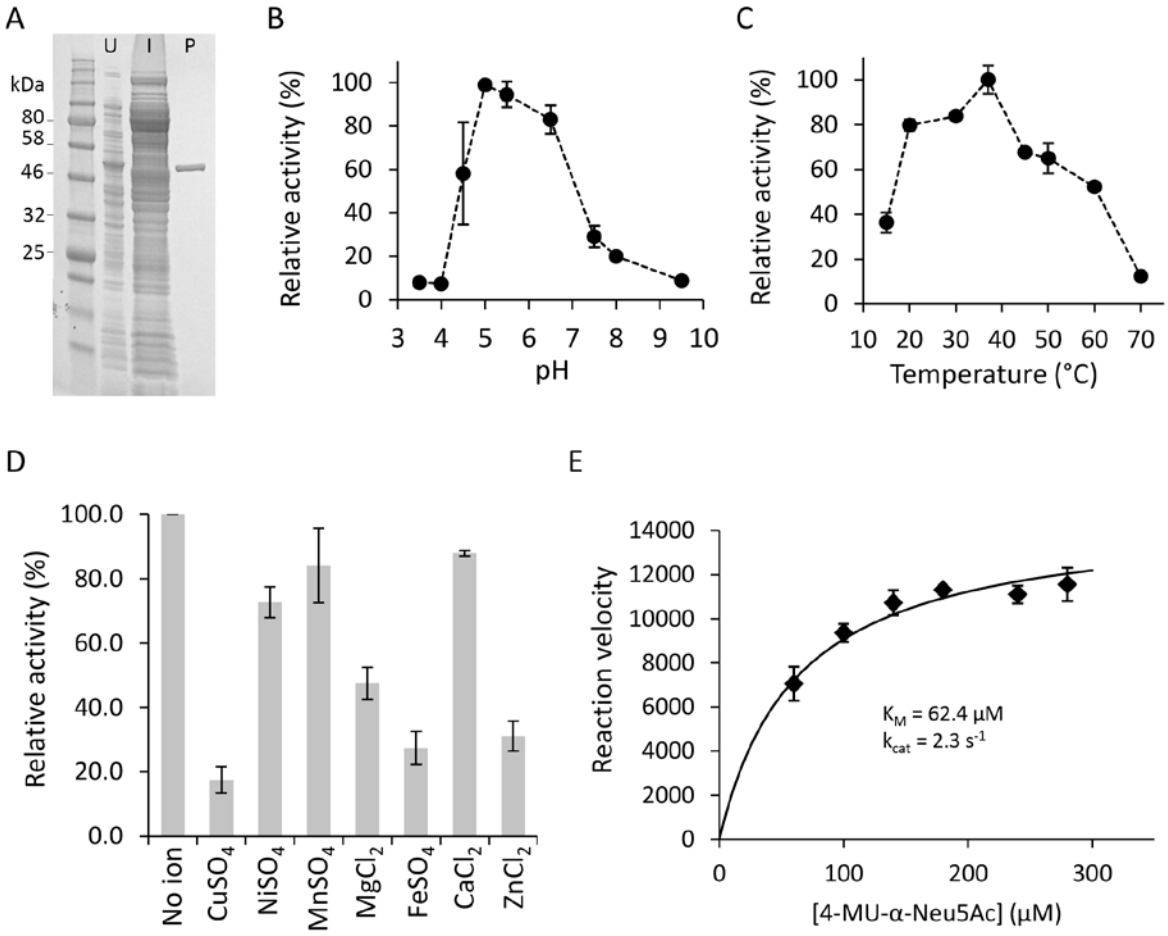


Figure 4. Purification and biochemical characterization of recombinant ORF12p. (A) His-tagged ORF12p sialidase was expressed in *E. coli* and purified using a His-trap column as described in the Experimental procedures. Shown is SDS-PAGE separation of lysates from uninduced cells (U), induced cells (I) and nickel-purified ORF12p-His (P). Purified ORF12p-His was used to determine the pH (B) and temperature (C) optima of ORF12p-His, and the effect of metal ions on its catalysis (D). In these experiments, reactions were performed in triplicate using the substrate 3'-sialyl-*N*-acetyllactosamine-2AB. Reaction products were analyzed by UPLC-HILIC-FLR and quantitated by peak integration. (E) Michaelis-Menten plot of ORF12p catalyzed hydrolysis of 4MU-α-Neu5Ac. The initial velocity was determined in triplicate for each 4MU-α-Neu5Ac concentration.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

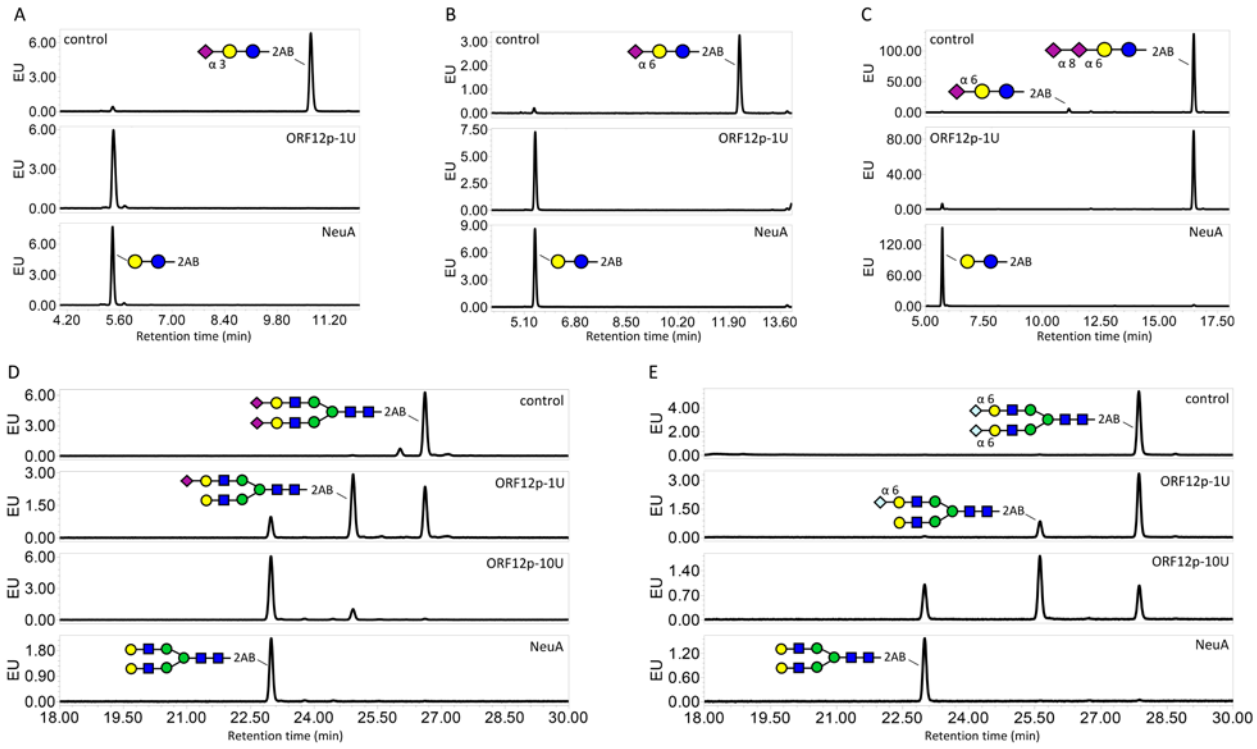


Figure 5. Specificity of ORF12p on sialic acid containing substrates using UPLC-HILIC-FLR analysis. (A, B) The ability of ORF12p to cleave the fluorescently labeled substrates 3' or 6'-sialyl-*N*-acetyllactosamine-2AB. Undigested substrates 3' or 6'-sialyllactosamine-2AB run at ~10.6 or 12.3 min retention times, respectively (A, B top panels). Control digestion with the NeuA sialidase shifted both substrate peaks to ~5.5 min retention time (A, B bottom panels). Digestion of these substrates with 1U of ORF12p-His resulted in the same peak shift (A, B middle panels). (C) ORF12p's ability to hydrolyze α2-8 Neu5Ac was assessed using 2-AB labeled a GD3 ganglioside head group substrate that contains two sialic acid residues linked via an α2-8 bond. Undigested substrate ran at ~16.5 min retention time with a very minor peak at ~11 min retention time corresponding to partially degraded substrate comprised of a single α2-6 terminal sialic acid (C top panel). NeuA treated substrate shifted at ~5.5 min retention time (C bottom panel). Treatment with 1U ORF12p did not shift the major substrate peak (C middle panel). (D, E) Activity of ORF12p on biantennary complex *N*-glycans with terminal sialic acid residues (Neu5Ac, D or Neu5Gc, E). Undigested substrates run at ~26.6 and 27.9 min retention time, respectively (D, E top panels). NeuA treatment shifted both substrate peaks at ~23 min retention time (D, E bottom panels). Incubation of the substrates with 1 or 10U of ORF12p resulted in the same peak shift but incomplete substrate desialylation was observed resulting in another smaller peak shift at ~24.9 and 25.6 min retention time, respectively (D, E middle panels). Symbolic representation of glycan structures was drawn following the guidelines of the Consortium for Functional Glycomics (50). Abbreviation: EU, emission units.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

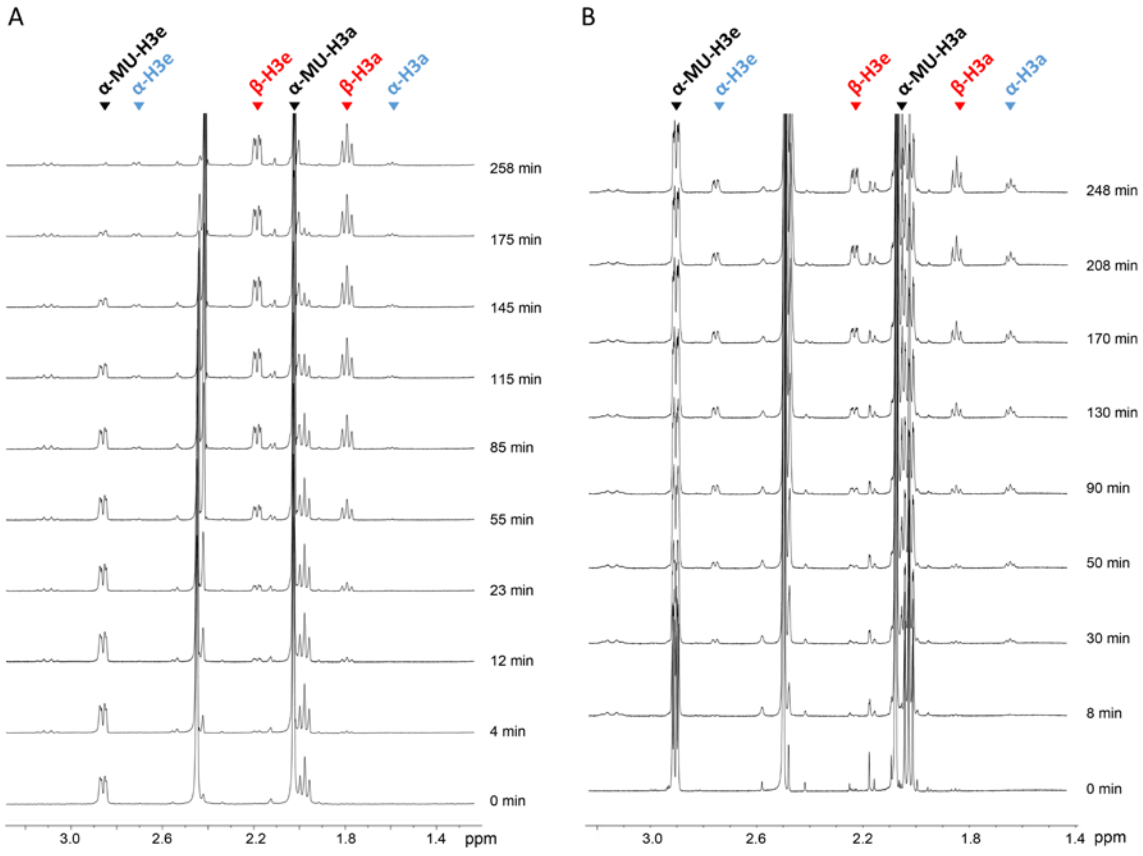


Figure 6. Stereochemical course of the hydrolysis reaction catalyzed by ORF12p (A) compared to NeuA (B). 1D proton NMR was used to monitor the reaction products formed over time upon hydrolysis of 4MU-α-Neu5Ac by ORF12p and NeuA. (A) The up-field region of the NMR spectra showed two groups of peaks at 1.99 and 2.86 ppm corresponding to the substrate 4MU-α-Neu5Ac (black triangles). Over time, groups of peaks appeared at 1.81 and 2.20 ppm corresponding to released β-Neu5Ac (red triangles). (B) The spectra showed the substrate 4MU-α-Neu5Ac groups of peaks at 2.03 and 2.89 ppm (black triangles). After 30 minutes, H3a and H3e peaks corresponding to the α-anomer appeared at 1.64 and 2.76 ppm (blue triangles). These were quickly converted by mutarotation to the β-anomer at 1.85 and 2.23 ppm (red triangles).

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

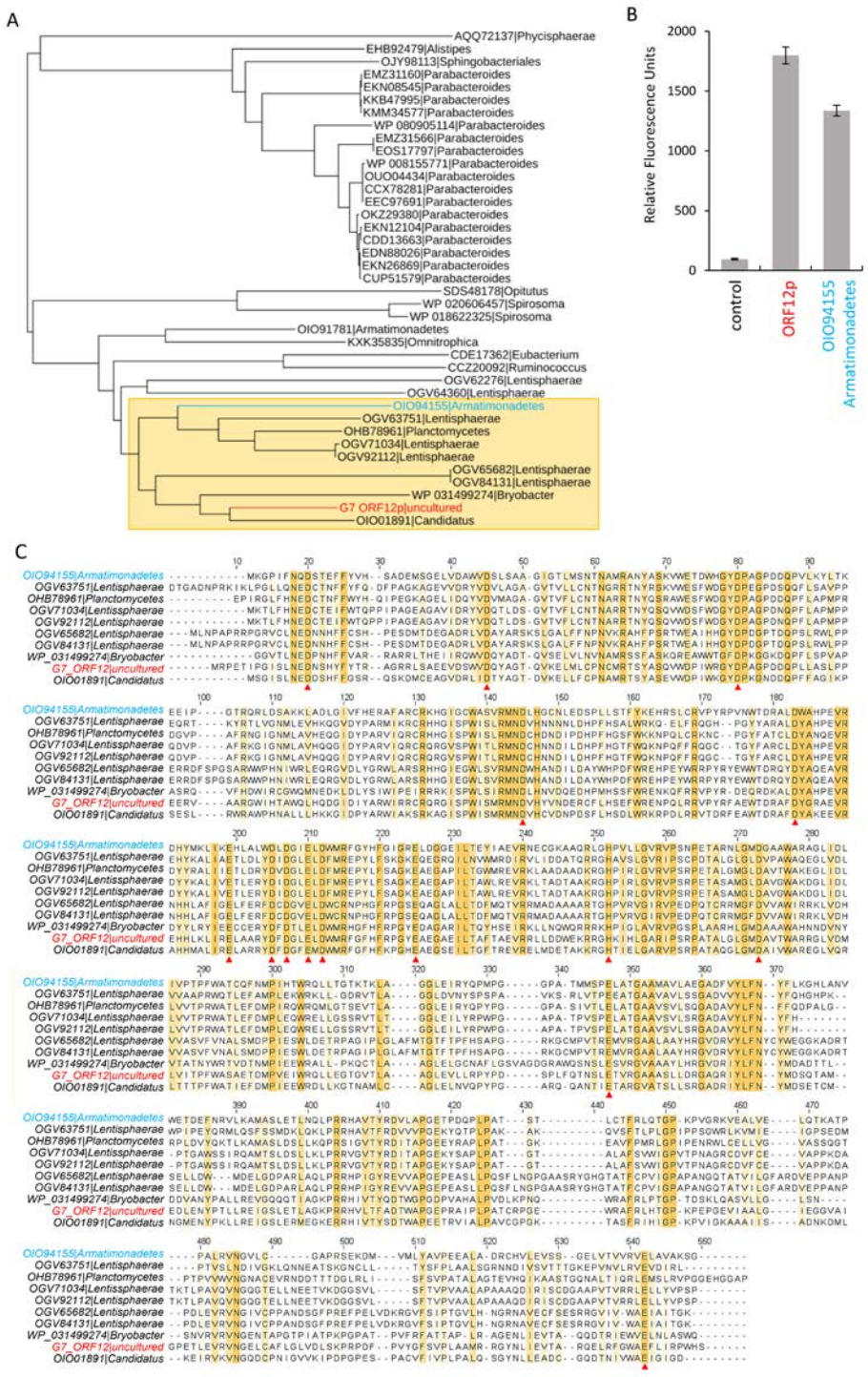


Figure 7. ORF12 protein family. (A) ORF12p homologs were identified with the BLASTP algorithm and aligned with Muscle. A phylogenetic tree was then generated with the BLOSSUM62 matrix and the Neighbor-Joining method. ORF12p and *Armatimonadetes* OIO94155 are shown in red and blue text, respectively. (B) The *Armatimonadetes* OIO94155 protein homolog was expressed in *E. coli* and crude lysate was tested with 4MU-α-Neu5Ac. (C) Protein sequence alignment (Muscle) of ORF12p and its closest homologs (A, yellow box), including *Armatimonadetes* OIO94155. Conserved residues are highlighted in yellow and those potentially involved in catalysis are shown with red triangles.

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020

Functional metagenomics identifies an exosialidase with an inverting catalytic mechanism that defines a new glycoside hydrolase family (GH156)

Léa Chuzel, Mehul B Ganatra, Erdmann Rapp, Bernard Henrissat and Christopher H Taron

J. Biol. Chem. published online September 24, 2018

Access the most updated version of this article at doi: 10.1074/jbc.RA118.003302

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Downloaded from http://www.jbc.org/ at DOE/SC/OSTI on May 22, 2020