# A novel enzyme isolated from a hot spring metagenomic DNA library defines a new CAZy family (GH154) of exosialidases with an inverting catalytic mechanism

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

[1]New England Biolabs, Ipswich, USA; [2]Max Planck Institute, Magdeburg, Germany; [3]CNRS, Aix-Marseille Université, France

June 2018

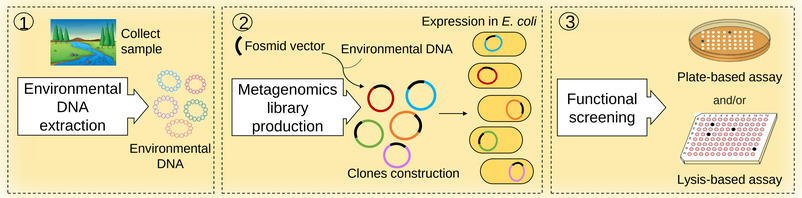
## ABSTRACT

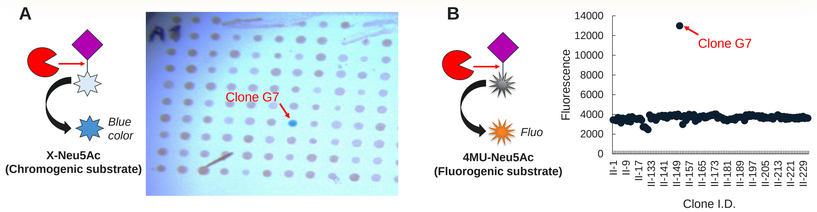
**Exosialidases** (also termed neuraminidases; E.C. 3.2.1.18) are glycoside hydrolases that catalyze removal of a single terminal sialic acid from a subterminal sugar in an oligosaccharide. They are widely distributed in biology, having been found in prokaryotes, eukaryotes and certain viruses. Most characterized prokaryotic sialidases derive from organisms that are pathogenic or commensal with mammals. Less is known about sialidases from non-commensal microorganisms, including those that thrive in an extreme environmental niche like hypersaline ponds, evaporation salterns or thermal springs. In this study, we sought to explore if active sialidases could be identified from organisms that populate a thermal spring.

To address this question, we constructed a **fosmid library *in Escherichia coli* from metagenomic DNA** that had been isolated from green microbial mats from the Dixie Hot Spring in Nevada. A total of 616 *E. coli* clones, each having a fosmid with an insert of ~40 kb of environmental DNA, were created and arrayed in microtiter plates for screening. The library was screened for sialidases with two substrates: 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid (4MU-Neu5Ac) and 5-bromo-4-chloro-3-indolyl α-D-N-acetylneuraminic acid (X-Neu5Ac).

**A single *E. coli* clone having sialidase activity** was identified using both substrates. The fosmid was isolated from this strain, sequenced using the Pacific Biosciences DNA sequencing platform, and encoded ORFs were predicted with MetaGeneMark. The DNA sequence did not match any reported sequences from known microorganisms. Additionally, none of the predicted ORFs showed homology to existing sialidase families. Tn5 mutagenesis was conducted to identify the **505 amino acid ORF** responsible for the enzymatic activity. BLASTP using the ORF's deduced protein sequence analysis indicated that it was a member of **a small family of bacterial "hypothetical" proteins with no known function**. The protein was recombinantly over-expressed in *E. coli* and was shown to hydrolyze a variety of sialic acid containing substrates. Additionally, protein NMR showed that the enzyme functions via an **inverting catalytic mechanism**, a biochemical property distinct from known exosialidases that each function via a retaining mechanism. This unique inverting exosialidase defines a **novel CAZy glycoside hydrolase family** that has been designated **GH154**.

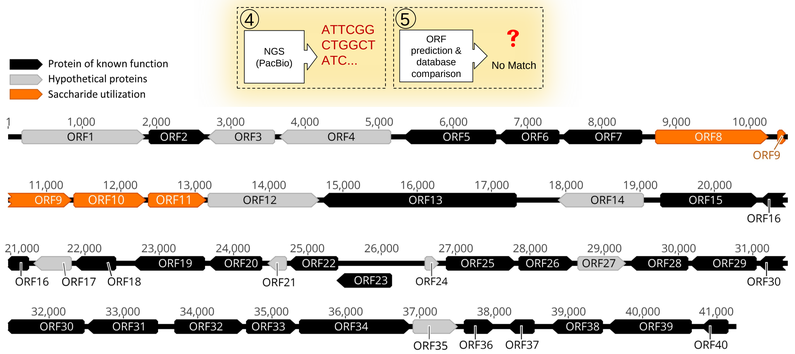
## FUNCTIONAL SCREENING



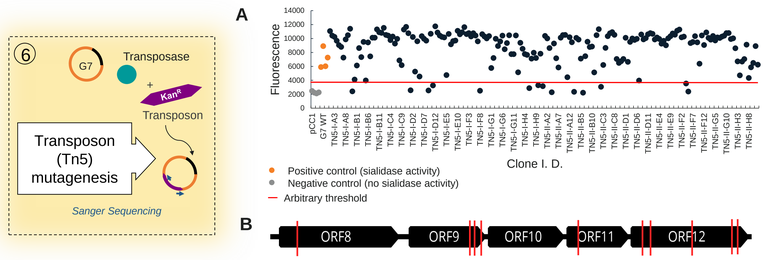

- A fosmid library (616 clones) containing metagenomic DNA from a hot spring was used for sialidase screening
- Sialidase activity was identified using a chromogenic substrate (**A**) and a fluorogenic substrate (**B**). A single clone (termed G7) was identified in both assays

## FOSMID G7 SEQUENCING



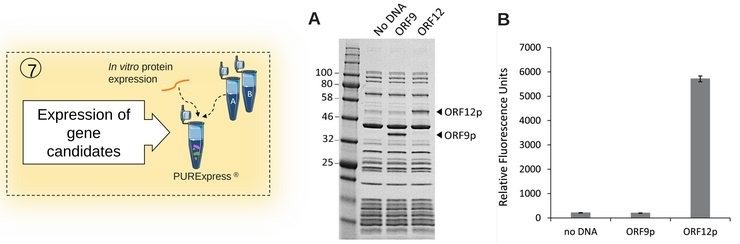

- Fosmid G7 was isolated and sequenced via PacBio RSII
- 40 different ORFs were identified using MetaGeneMark
- BLAST analysis revealed no homology with known sialidase CAZy families (GH33, GH34, GH58 and GH83)

## Tn5 MUTAGENESIS



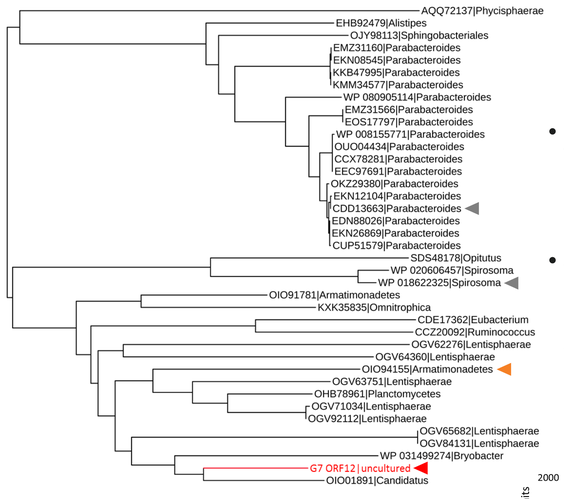

- A kanamycin cassette was randomly inserted into fosmid G7 using Tn5 transposon mutagenesis
- 196 Tn5 transposon mutants were screened for sialidase activity (4MU-Neu5Ac)
- Clones with reduced or abolished sialidase activity (**A**, below red line) were Sanger sequenced
- Tn5 integrations clustered to ORF9 and ORF12 (**B**, red line)

## SIALIDASE IDENTIFICATION

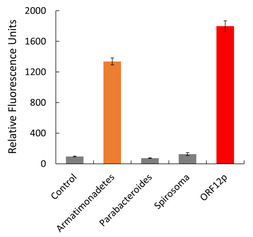



- ORFs 9 and 12 were expressed *in vitro* using PURExpress (**A**, black arrows)
- Reaction products were assayed for sialidase activity (**B**, 4MU-Neu5Ac)
- Only the translated product of ORF12 (ORF12p) showed sialidase activity (**B**)

## ORF12p FAMILY

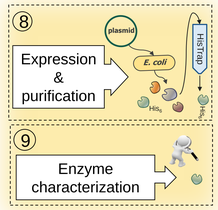



- ORF12p (red) homologs were identified with the BLASTP algorithm and aligned with Muscle
- A phylogenetic tree was generated (BLOSSUM62 matrix and the Neighbor-Joining method)

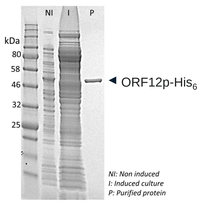



- 3 homologs of ORF12p were expressed in *E. coli*
- Crude lysate activity was tested on 4MU-Neu5Ac
- *Armatimonadetes* OIO94155 protein showed sialidase activity (orange)

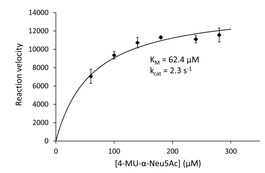
## ORF12p PURIFICATION & CHARACTERIZATION




- ORF12p *E. coli* expression



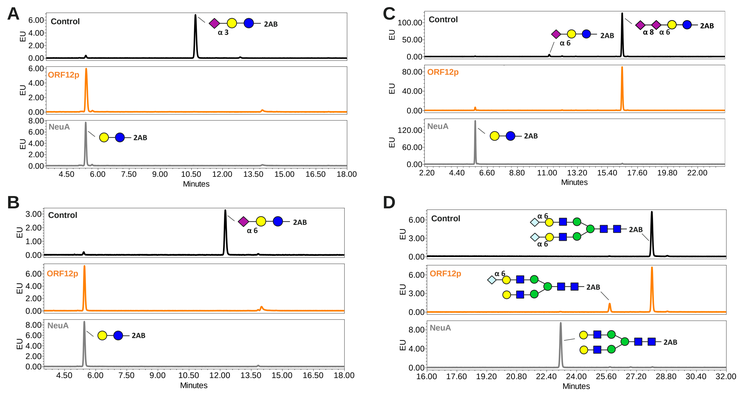

- Michaelis-Menten plot of ORF12p

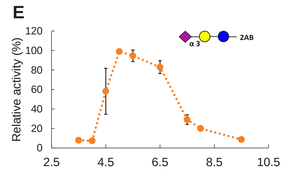


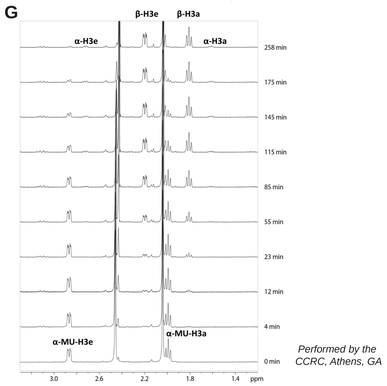

- Specificity of purified ORF12p was assessed with:

(A) α3 2AB (C) α8 α6 2AB
(B) α6 2AB (D) α6 α6 2AB

- Reaction analysis by UPLC-HILIC-FLD




- The pH (**E**) and temperature (**F**) optima of ORF12p were determined
- Reaction analysis by UPLC-HILIC-FLD
- Peak areas quantitated by integration
- The reaction mechanism of ORF12p was assessed using 1D proton NMR (**G**)
- 4MU-α-Neu5Ac was used as the substrate
- Over 258 minutes, β-Neu5Ac was released
- ORF12p functions via an inverting mechanism

## CONCLUSION

- We have discovered a **novel exosialidase** using functional metagenomics
- The enzyme shows no sequence homology to known sialidases or other proteins of known function and is part of a family of >30 bacterial proteins of previously unknown function
- The enzyme hydrolyzes terminal α 2,3- and α 2,6-linked Neu5Ac via an **inverting mechanism**
- This enzyme represents a **new CAZy family** designated **GH154**