Characterization of α2,3- and α2,6-sialyltransferases from Helicobacter acinonychis

Melissa J Schur2, Emilie Lameignere3, Natalie CJ Strynadka3, and Warren W Wakarchuk1,2,4
1National Research Council Canada, Institute for Biological Sciences, 100 Sussex Drive, Ottawa, ON, Canada K1A 0R6; 2Department of Biochemistry and Molecular Biology, Centre for Blood Research, University of British Columbia, Vancouver, BC, Canada V6T 1Z3; and 3Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada N1G 2W1

Received on March 9, 2012; revised on April 4, 2012; accepted on April 5, 2012

Genome sequence data were used to clone and express two sialyltransferase enzymes of the GT-42 family from Helicobacter acinonychis ATCC 51104, a gastric disease isolate from Cheetahs. The deposited genome sequence for these genes contains a large number of tandem repeat sequences in each of them: HAC1267 (RQKELE)15 and HAC1268 (EEKLLEFKNI)13. We obtained two clones with different numbers of repeat sequences for the HAC1267 gene homolog and a single clone for the HAC1268 gene homolog. Both genes could be expressed in Escherichia coli and sialyltransferase activity was measured using synthetic acceptor substrates containing a variety of terminal sugars. Both enzymes were shown to have a preference for N-acetyllactosamine, and they each made a product with a different linkage to the terminal galactose. HAC1267 is a mono-functional α2,3-sialyltransferase, whereas HAC1268 is a mono-functional α2,6-sialyltransferase and is the first member of GT-42 to show α2,6-sialyltransferase activity.

Keywords: glycosyltransferase / Helicobacter / sialyltransferase

Introduction

Bacterial pathogens that colonize mucosal surfaces make use of their surface carbohydrate structures either to mask themselves from immune system detection (Guerry et al. 2000; Gulati et al. 2005) or to bind to host receptors to aid in colonization or transmission of the pathogen (Harvey et al. 2000; Carlin et al. 2009). These bacterial carbohydrate structures that mimic host glycans can be capsular polysaccharide or lipopolysaccharides (LPSs). In the case of LPSs, they are often the short-chain variants that are lacking the O-antigen repeating unit (lipooligosaccharide, LOS) and often mimic host glycolipid carbohydrate structures, including blood group antigens and gangliosides (Gilbert et al. 2008; Houliston et al. 2011). Many bacterial pathogens also regulate the composition of the terminal LPS glycan through a variety of genetic means. In particular, Campylobacter jejuni, Helicobacter pylori and Haemophilus influenzae achieve a large number of possible LOS structures using genes whose expression is phase variable (Schweda et al. 2007; Gilbert et al. 2008; Nilsson et al. 2008).

The rapid determination of bacterial pathogen genome sequences has increased the number of members of many virulence-associated, LPS glycosyltransferases. Based on the automated annotation of genes, it appears that many bacteria possess similar glycosyltransferase activities but very few of these have been biochemically verified. Since there are many gene variations that can give rise to different glycosyltransferase activities (Gilbert et al. 2008), simply having the sequence does not permit accurate prediction of resulting LPS structures. We have been characterizing bacterial sialyltransferases involved in LPS/LOS biosynthesis and have previously investigated the structure and function of the enzymes in the CAZY (Cantarel et al. 2009) GT-42 family (Gilbert et al. 2002; Chiu, Watts, et al. 2004; Chiu, Lairson, et al. 2007; Lee et al. 2011). When the biochemical work on the GT-42 sialyltransferases started, there were corresponding data on the cell surface LPS structures for many C. jejuni isolates, which guided the design of assays to detect and characterize the sialyltransferases (Aspinall, Fujimoto, et al. 1994; Aspinall, McDonald, et al. 1994; Penner and Aspinall 1997). The glycosyltransferase family GT-42 has a number of members that have been characterized as mono-functional α2,3,3-sialyltransferases as well as the unusual bi-functional α2,3/α2,8-sialyltransferases (Gilbert et al. 2000; Fox et al. 2006; Thon et al. 2011). Using the sequence and structural information of the GT-42 family members, it has not yet been possible to accurately predict which enzymes will be mono- or bi-functional except within the large group of closely related enzymes from C. jejuni (Gilbert et al. 2002).

We have been investigating enzymes from the GT-42 family from recently sequenced bacteria for which there may not be corresponding data about the LPS structure. We cloned and expressed two putative GT-42 genes from H. bizzozeronii str. Storkis and were able to show that only one of these, HBS-02, was a functional sialyltransferase based on the

To whom correspondence should be addressed: Tel: +1-613-952-4299; Fax: +1-613-941-1327; e-mail: warren.wakarchuk@nrc-cnrc.gc.ca
activity of the recombinant protein (Kondadi et al. in press). The second gene appears to code for an inactive protein, possibly because of critical amino acid substitutions close to known active site residues, or other insertions/deletions that are deleterious to the folding or activity of the protein. A sequence alignment of these representative GT-42 enzymes which have been examined for activity is shown in Figure 1.

*Helicobacter* strains show a variety of mechanisms for their phase variable LPS glycosyltransferases. Fucosyltransferases from *Helicobacter* show the phase variability of either through homopolymeric tracts or repeated sequences that can lead to slip-strand mispairing during DNA replication (Nilsson et al. 2006; Nilsson et al. 2008). The *Helicobacter* fucosyltransferases of GT-10 show internal repeat sequences of seven amino acids which are variable and may function as an environmental sensor for the pathogen expression of fucosylated Lewis structures (Skoglund et al. 2009). There is a related amino acid repeat in the GT-42 sialyltransferases in *Helicobacter acinonychis* str. Sheeba and some *H. pylori* isolates (Figure 2). It is not yet known if these sequences are involved in the phase variable expression of sialylated LPS, but it is likely that they are based on the data from the GT-10 fucosyltransferases.

In this paper, we examine the recombinant expression and characterization of the two GT-42 genes found in *H. acinonychis* ATCC 51104 and we show that one of these, HAC1268, is an α2,6-sialyltransferase, the first such activity shown in CAZy family GT-42.

**Results**

**Cloning of the genes based on deposited sequences for HAC1267 and HAC1268**

We were able to amplify both genes from the genomic DNA of *H. acinonychis* ATCC 51104 based on the genome sequence for *H. acinonychis* str. Sheeba (GeneBank NC_008229.1 GI:109946640). We obtained three distinct clones for the HAC1267 gene all with four nucleotide changes from the deposited gene sequence, which are silent changes in the protein sequence; one clone was full length, one was missing 4 of the 15 repeats of the 6-amino acid sequence RQKELE and one, missing five of the six-amino acid sequence repeats (Figure 2). For the HAC1268 gene, we only obtained clones with a single sequence from our isolated genomic DNA, and it was missing 11 of the 13 amino acids repeats of the 10-amino acid sequence EEKLLEFKNI, as well as having 18 nucleotide changes from the deposited gene sequence. The 18 nucleotide changes in HAC1268 lead to three-amino acid substitutions from the deposited genome, Ser61Pro, Val66Ala and Ser175His. The original amplicons of HAC1267 showed
a mixture of two major amplicons (data not shown), but the reaction for HAC1268 showed only one clean PCR product corresponding to the truncated gene we cloned.

Expression of sialyltransferase activity and optimization of the gene constructs

The initial clones were expressed in the expression plasmid pCWori+ and all of these clones expressed small amounts of enzyme activity detected using the synthetic substrates FEX (or FCHASE)-LacNAc. Since there was no appreciable difference in activity between the full-length and truncated HAC1267 proteins, we only worked further with the truncated form. Previous work with the *H. bizzozeronii* HBS-02 GT-42 sialyltransferase had shown that an increase in expression could be obtained by making fusions to the *Escherichia coli* maltose-binding protein (MalE). We constructed MalE fusions of both proteins and were able to see an increase in the amount of protein being produced (data not shown), although not a substantial increase in the amount of enzyme activity. For the HAC1267, we saw an increase in the enzyme activity of \( \approx 6 \)-fold for the shorter variant relative to the full-length clone (Table I). The amount of active enzyme for the HAC1268 enzyme was not improved by making the MalE fusion. Sequence comparisons between various members of GT-42 and previous experience with the expression of this class of transferases suggested that we could truncate HAC1268 by 16 amino acids to make its C terminus a similar length to that of HAC1267 (Figure 1). The C-terminal \( \Delta 16 \) amino acid construct showed a 17-fold improvement in active protein production over the initial MalE fusion protein and 3-fold better than the original clone (Table I).
For the expression of many bacterial sialyltransferases in *E. coli*, we have used the strain AD202 (CGSC 7297) and had very good expression and recovery of active proteins, but we have on occasion tried other *E. coli* strains to improve the production of active enzyme. We found a further improvement in the production of active protein by expressing the HAC1267/1268-MalE fusions in the SHuffle™ Express strain from New England Biolabs, Mississauga, ON. We saw a further 8-fold increase in the amount of active HAC1267 protein and a further 2.5-fold increase in the amount of HAC1268 protein (Table I).

### Characterization of the sialyltransferases from HAC1267 and HAC1268

The MalE fusions of the two proteins could be purified by affinity chromatography on amylose resin. The activity of the proteins was examined under a variety of pH (data not shown), with and without divalent metals and with a number of synthetic acceptor substrates. The data are summarized in Figure 3 and Table II. Both of these enzymes show a strong preference for N-acetyllactosamine which has been shown to be present in the LPS from *H. pylori* (Monteiro 2001) and recently in *H. bizzozeronii* (Kondadi et al. in press). Like the other members of the GT-42 family, these enzymes have no absolute requirement for divalent metal ions. HAC1267 shows some good improvement in activity with Mg+2 and a slight decrease in activity in the presence of Mn+2, whereas the situation is different with HAC1268 (Figure 3). With HAC1268, the addition of 10 mM Mn+2 decreases the activity and seems to have a progressive toxic effect in longer reactions with the rate decreasing. The addition of Mg+2 has a small beneficial effect, and the addition of EDTA has no real effect. The HAC1268 enzyme can be protected from Mn+2 poisoning by pre-incubating with DTT, and in fact pre-incubation with DTT leads to slightly higher enzyme activity. Control reactions with HAC1268 got to 30% conversion in 45 min, with 10 mM MnCl2 only 16% conversion is reached, with DTT added to the reaction they reach 75% conversion, a pre-incubation with DTT has a more immediate effect and also reaches ~75% conversion and with DTT pre-incubation and then 10 mM MnCl2 in the reaction gets to 60% conversion.

The acceptor specificity of GT-42 enzymes has been examined for many family members. We showed previously with several of the GT-42 proteins from *C. jejuni* that these enzymes use both type I and II linkages equally well (Chiu, Lairson, et al. 2007). In all, we screened HAC1267/1268 against fluorescent derivatives of Lac, LacNAc, Lacto-N-biose, Gal-β1,3-GalNAc, β- and α-GalNAc, βGal, βGlcNAc, Lewis X and 3'-sialyl-Lac (GM3) as acceptors. Only Lac, LacNAc and Lacto-N-biose showed acceptor activity, with only a trace of activity being seen for Gal-β1,3-GalNAc. HAC1267 and HAC1268 were ~10-fold more active on LacNAc than Lac (Table I).

When we assayed these enzymes and analyzed the products by either TLC or capillary electrophoresis (CE), we noticed a shift in the mobility of the product from HAC1268 reactions which suggested a different linkage to the terminal galactose of those acceptors (data not shown). HAC1268 reaction products co-migrated with known α2,3-linked sialylated products (data not shown), whereas products from HAC1268 reactions were different. The NMR analysis of the purified reaction product from an HAC1268 reaction on FEX-LacNAc showed that the linkage was indeed α2,6 to the terminal galactose (Supplementary data, Table SI) with chemical shifts almost identical to the similar structure found in the *Streptococcus suis* serogroup 2 CPS (Van Calsteren et al. 2010).

### Comparison of HAC1267 and HAC1268 with other GT-42 family members

Enzyme activity. Detailed structure–function studies have been performed on the *C. jejuni* Cst-II protein (Chiu, Watts, et al. 2004; Chiu, Lairson, et al. 2007). The first three constructs were expressed in *E. coli* AD202. Activity was compared after re-suspending cells at the same relative concentration before making the extract.

### Table I. Increases in enzyme activity as a consequence of protein engineering and host strain variation

<table>
<thead>
<tr>
<th>Construct</th>
<th>Activity (mU/mL)</th>
<th>Fold increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-length gene</td>
<td>3.7</td>
<td>—</td>
</tr>
<tr>
<td>Missing four repeats (HAC1267 d4)</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>MalE: HAC1267 d4</td>
<td>20</td>
<td>5.4</td>
</tr>
<tr>
<td>Expressed in Shuffle™</td>
<td>179</td>
<td>48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct</th>
<th>Activity (mU/mL)</th>
<th>Fold increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing 11 repeats (HAC1268 d11)</td>
<td>10.3</td>
<td>—</td>
</tr>
<tr>
<td>MalE: HAC1268 d11</td>
<td>2.2</td>
<td>—4.6</td>
</tr>
<tr>
<td>MalE: HAC1268 d11 Δ16 C-terminal aa</td>
<td>38</td>
<td>3.7</td>
</tr>
<tr>
<td>Expressed in Shuffle™</td>
<td>93.5</td>
<td>9.1</td>
</tr>
</tbody>
</table>

The first three constructs were expressed in *E. coli* AD202.
et al. 2004; Lee et al. 2011), and these data can be used to examine a sequence alignment of other GT-42 family members. A protein sequence alignment of GT-42s from C. jejuni (Cst-I and Cst-II), H. bizzozeronii (HBS-01 and HBS-02), H. acinonychis (HAC1267 and HAC1268) and Pasteurella multocida PM1174 is shown in Figure 1. In order to better understand the differences in enzyme activity between these members of the GT-42 family, we examined all these proteins and performed a comparison GT-42 enzyme from C. jejuni, H. bizzozeronii, H. acinonychis and P. multocida with a variety of acceptors having both type I and II linkages. These results are shown in Table II. These assays were performed using the optimized conditions for each of the enzymes, using a standard assay containing 0.5 mM acceptor and 1.0 mM donor. The three Helicobacter enzymes all prefer LacNac as an acceptor and have a rather narrow acceptor specificity. The activity of the C. jejuni enzymes is much higher than any of the other enzymes and of a much broader specificity.

Structural models. There are two structures of GT-42 enzymes which we used for molecular modeling of the HAC1267/1268 enzymes to provide additional insight into their activity. Both the HAC1267 and HAC1268 proteins were built with more than 90% confidence. The final model of the HAC1268 protein was built at 99%, whereas the HAC1267 model is complete at 100%. The HAC1268 protein model was generated using the closest two homologs of known structure: CstII PDB-ID: 1RO7 and Cst PDB-ID: 2P2V, 2WQQ as templates. Sixty-seven residues including the first four amino acids at the N-terminal region, the C-terminal part (260–312) and an extra loop (90–98) close to the acceptor binding site of the protein were built ab initio, resulting in a low-confidence key for those regions. The HAC1267 was built using the same three templates as for HAC1268. In addition, the C-terminal domain (210–371) was built using the MBP protein (PDB-ID: 1y4c) and the rat liver vault (PDB-ID: 2xv4). Only one amino acid was built ab initio. Only the highly conserved N-terminal segments of proteins (residues 4–220 for HAC1267 and 6–258 for the HAC1268), which were modeled using Cst and CstII as templates, are presented here. The two models reveal a highly similar overall architecture to the CstI and CstII monomer structures (Figures 4 and 5).

Structure–function studies with CstI and CstII have identified key residues involved in substrate binding and catalysis. If only the primary sequence alignment is used, then the prediction would be that some known active site residues are conserved in HAC1268 (Figure 1). A summary of key residues observed in the CstII structure compared with those observed in the HAC1267/1268 models is shown in Table III. These residues are those involved in stabilizing the phosphate moieties of CMP in the Cst proteins (Asn31, Tyr156 and Tyr162 in CstII numbering), only the Asn31 is conserved in both HAC1267 and HAC1268. The
structure alignment shows that in the HAC proteins, His (H167 in HAC1267 and H172 in HAC1268) could bind to the phosphate moiety of CMP. According to the model, the following Tyr (Y169 and Y174) are too far away to stabilize the phosphate moiety and play the same role as the Tyr162 (Figure 4). Both HAC enzymes lack key residues seen in the bi-functional CstII enzyme which is consistent with their mono-functional transferase activity.

Helicobacter GT-42 with repeats assort into two distinct groups

When all of the CAZy GT-42 Helicobacter sequences with repeat sequences are aligned, they clearly form two groups. These groups are those very similar to HAC1267 (four with >90% identity) and those that are very similar to HAC1268 (two with >95% identity). Based on our activity assays with HAC1267 and HAC1268, we can predict that these groups share the enzymatic activity of the HAC enzymes.

Discussion

The presence of sialic acids in glycoconjugates from bacteria which mimic human carbohydrate epitopes is common in mucosal pathogens, and glycoconjugate structures have been characterized in a number of organisms that infect humans (Carlin et al. 2009; Virji 2009; Schweda et al. 2007; Houliston et al. 2011; Kondadi et al. in press). Examination of the biosynthetic pathways for these molecular mimic structures has led to the characterization of several sialyltransferase enzymes. Sialyltransferases from bacterial species make up the CAZy families GT-38, GT-42, GT-52 and GT-80. Of these families, only GT-42 and GT-52 enzymes have been shown to participate in adding sialic acid to LPSs in the parent organisms. Enzymes from GT-52 are found in Neisseria, Haemophilus and Pasteurella species, and a role in LPS biosynthesis has been shown only in Neisseria and Haemophilus species (Gilbert et al. 1996; Jones et al. 2002). So far there is no evidence that Pasteurella species produce sialylated LPS (St. Michael et al. 2005), despite the presence of three sialyltransferases in deposited genome sequences.

The sialyltransferases from GT-42 were the first sialyltransferases for which a structure was available (Chiu, Watts, et al. 2004), and they were first characterized from Campylobacter and later from Haemophilus (Gilbert et al. 2000; Fox et al. 2006). Some members of this family were shown to be capable of bi-functional reactions where there was an \( \alpha_2,3 \)-sialyltransferase activity and a second \( \alpha_2,8 \)-sialyltransferase activity. The structure and function of the GT-42 family have been extensively characterized, and there are multiple protein structures of various forms of the

Table III. Amino acid involved in substrate interaction and catalysis in GT-42 enzymes

<table>
<thead>
<tr>
<th>CSTI</th>
<th>CSTII</th>
<th>HAC1267</th>
<th>HAC1268</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermolecular contacts stabilizing the phosphate moiety of CMP in the Cst proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn46</td>
<td>Asn31</td>
<td>Asn33</td>
<td>Asn35</td>
</tr>
<tr>
<td>Tyr171</td>
<td>Tyr156</td>
<td>Tyr161</td>
<td>Gly166</td>
</tr>
<tr>
<td>Tyr177</td>
<td>Tyr162</td>
<td>His167</td>
<td>His172</td>
</tr>
<tr>
<td>Specific amino acids for bi-functional Cst-II variants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn66</td>
<td>Asn51</td>
<td>Asn53</td>
<td>Asn55</td>
</tr>
<tr>
<td>Val69</td>
<td>Leu54</td>
<td>Val56</td>
<td>Val58</td>
</tr>
<tr>
<td>Phe284</td>
<td>Ile269</td>
<td>Lys256</td>
<td>Ser242</td>
</tr>
<tr>
<td>Catalytic histidine</td>
<td>His188</td>
<td>His184</td>
<td>His190</td>
</tr>
<tr>
<td>Acceptor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galβ1-3/4R</td>
<td>Galβ1-3/4R</td>
<td>LacNAc&gt;Lac</td>
<td>LacNAc&gt;Lac</td>
</tr>
<tr>
<td>Neu5Aco2-3Galβ1-3/4R</td>
<td></td>
<td>&gt;&gt;Galβ1-3GlcNac (trace)</td>
<td>&gt;&gt;Galβ1-3GlcNac (trace)</td>
</tr>
<tr>
<td>Linkage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mono-functional ( \alpha_2-3 )-sialylated</td>
<td>Bi-functional ( \alpha_2-3/\alpha_2-8 )-sialylated</td>
<td>Mono-functional ( \alpha_2-3 )-sialylated</td>
<td>Mono-functional ( \alpha_2-6 )-sialylated</td>
</tr>
</tbody>
</table>

Fig. 5. Structural alignment of the CstI/II overall structure with the HAC1267 and HAC1268 models. The cartoon trace of CstI (PDB2X61) is shown in grey. The stick representation of the donor (CMP) and trisaccharide acceptor (Neu5Ac-\( \alpha \)-2,3-Gal-\( \beta \)-1,3-GalNac) are shown in yellow and blue, respectively; non-carbon atoms are colored according to type (nitrogen, blue; oxygen, red; phosphorus, orange). HAC1267 partial model (4–220) and HAC1268 partial model (6–258) are shown in green and violet. The HAC1268 loop (90–98) built \textit{ab initio} is represented in red.
bases of GT-42 enzymes in Helicobacter species. Until recently, there was no structural information on the LPS from these species to indicate a sialylated glycoconjugate. We recently reported that the H. bizzozeronii species contains sialyl-LacNAc in the LPS (Kondadi et al. in press). The presence of this carbohydrate epitope may have played a role in the transmission of this strain to humans where it now causes disease. We have not yet shown that the H. acinonychis LPS contains sialic acid, but the phase variation of LPS in Helicobacter strains is very common, and perhaps the isolate we obtained has another transferase turned off leading to a lack of an acceptor for the GT-42 enzymes. The LacNAc disaccharide, Lewis X and sialyl-Lewis X have all been shown in certain H. pylori strains based on the mass spectral analysis of LPS fragments (Monteiro et al. 2000). Many of the Helicobacter species with GT-42 enzymes have two copies of the gene, much as we observed previously in H. influenzae. The duplicate GT-42 enzymes in H. influenzae encode mono- and bi-functional versions of the enzyme (Fox et al. 2006). Our examination of the H. bizzozeronii GT-42 showed that only one enzyme had sialyltransferase activity on synthetic acceptors (Kondadi et al. in press).

Examination of the two H. acinonychis GT-42 genes reveals one to be a new α2,6-sialyltransferase and the other to be an α2,3-sialyltransferase with similar activity to the one we characterized from H. bizzozeronii. The Helicobacter GT-42 subfamily has three members which are likely to be α2,6-sialyltransferases as they share very high-sequence homology with the HAC1268 enzyme we characterized. The HAC1268 enzyme we isolated does have three amino acid changes relative to the deposited sequence and to the other two H. pylori “HAC1268-like” proteins, but these residues do not occur in the active site and so may not be important for the unique specificity of this enzyme. Of the sialic acid containing bacterial glycoconjugates that have been characterized, the presence of an α2,6-sialic acid is rare compared with the α2,3-sialylated structures, and so far has only been shown in LPS of Neisseria (Wakarchuk et al. 1998) and in the capsule of S. suis (Van Calsteren et al. 2010). Several α2,6-sialyltransferases from the CAZy GT-80 family produced by the organisms Photobacterium and Pasteurella have been characterized but there is no information for the sialylated glycoconjugate produced by these organisms (Yu et al. 2005; Kakuta et al. 2008). The only other bacterial α2,6-sialyltransferase to be characterized is from Neisseria meningitidis and is a GT-52 family member with a unique GT-B fold (Lin et al. 2011) that has a unique transfer activity to an α-linked Gal residue in the LPS.

We performed a comparison of six members of GT-42 with respect to their activity on a panel of synthetic acceptor substrates. The first observation we made was that the enzymes CstI and CstII are much more potent in vitro catalysts than any of the other enzymes we tested. The Helicobacter enzymes have a very strong preference for LacNAc, but they are 12–32-fold less active than CstI and 2.9-fold less active than CstII on that acceptor. CstI shows the highest level of promiscuity in that it is insensitive to 3- or 4-linked galactose residues. The other observation we made was that the HAC1267 and HAC1268 show improved recovery of enzyme activity when expressed in E. coli SHuffle Express™ cells. None of the other GT-42 enzymes we have examined show an improvement by expression in this strain. Molecular modeling of the HAC1268 enzyme does not suggest that any pairs of Cys would be close enough to form disulfides, so perhaps the increase in activity is simply due to a chaperone-like activity of the cytoplasmic DsbC protein in this strain (Chen et al. 1999).

The model of HAC1268 also provides possible reasons why the enzyme is poisoned by the addition of Mn²⁺ to reactions. There is Cys57 in the active site of the model structure which could be oxidized by Mn²⁺ and block binding of either acceptor or donor. A similar situation has been seen in the GT-8 galactosyltransferase LgtC, where Cys246 can hydrogen bond to the lactose acceptor and this enzyme requires DTT pre-incubation for maximum activity. The removal of Cys246 leads to the loss of the need for DTT to protect LgtC from oxidation (Persson et al. 2001).

We cannot tell from the model why the HAC1268 enzyme would make an α2,6-linked product. There is a possibility that the loop from residues 90–98 plays a role in binding the acceptor to achieve this specificity, but since the loop is only modelled at low confidence, this is only speculation. This question can really only be addressed through structure determination, and now that we can produce a reasonable quantity of active protein we are pursuing a more rigorous purification to obtain material for protein crystallography. It would be ideal to compare HAC1267 and HAC1268 together to answer this question.

We have now examined multiple members of the GT-42 family and have shown α2,3/2,6/2,8-sialyltransferase activities, which is similar to the eukaryotic sialyltransferase family GT-29 which also has all of these activities. This suggests that the GT-42 structure is flexible enough to permit the evolution of multiple transferase activities, as has happened with the GT-29 protein fold. This finding also serves to reiterate that membership in a glycosyltransferase family cannot accurately predict specificity until enough members have been assayed to permit conclusions about the enzymatic function. It also shows that these bacterial pathogens are well suited to employing a variety of terminal sialic acid containing structures to further their biological interactions with the host.

**Materials and methods**

**Bacterial strains, cultivation and DNA extraction**

The bacterial strains and plasmids used in this study are listed in Table IV. Helicobacter strains were routinely grown on HP medium (LabM Limited, Lancashire, UK) containing 5% bovine blood and Skirrow selective supplement (Oxoid Ltd,
Expression, purification and characterization of recombinant GT-42 enzymes

Primer pairs were designed to amplify complete sequences of HAC1267 and HAC1268 genes using Phusion High-Fidelity DNA Polymerase (NEB), and standard PCR conditions suggested by the manufacturer were used to perform the amplification (Table IV). The PM1174 gene was amplified from Pasteurella multocida PM70 genomic DNA and cloned in a similar fashion. The PM1174 protein was purified using the protocol developed for CstII (see below for reference). All genes were inserted in NdeI and SalI restriction sites of the expression vector pCwori+ (Wakarchuk et al. 1994). The resulting expression constructs were sequenced and confirmed to be HAC1267, HAC1268 and PM1174, respectively. The proteins CstI and CstII were expressed and purified as described previously (Chiu, Watts, et al. 2004; Chiu, Lairson, et al. 2007). All expression constructs were used to initially transform E. coli AD202 (Akiyama and Ito 1990). The genes were re-cloned as fusions with the MalE protein in the plasmid pCw-MaIE as described previously (Bernatchez et al. 2007). Expression of the MalE fusions in the E. coli SHuffle Express strain was performed in the same manner as for AD202. Expression of recombinant proteins was induced with 0.5 mM of one of the following acceptors: FCHASE-GM3, ampicillin and cloned in a similar fashion. The PM1174 protein was

---

**Table IV. Strains, plasmids and primers used in this work**

<table>
<thead>
<tr>
<th>Strains or plasmids</th>
<th>Relevant characteristics or genotype</th>
<th>References/comments</th>
</tr>
</thead>
</table>
| *Helicobacter* species
  *H. acinonychis* ATCC 51104 | HAC1267/HAC1268 GT42 | New England Biolabs |
| *Escherichia coli* stains
  Shuffle Express | fhuA2 [lon] ompT ahpC gal latt; pNEB3-r1-cDsBc (SpecR, lacIq) DrxA1 sulA11 R (merc-35: miniTn10-TetS2) [dcm] R(roz-h-201::Tn10–TetS) endA1 Agor Δ(mcrC-mrr)14::IS10 | Akiyama and Ito (1990) |
| AD202 | F' ompT::Km DE(argFlac)U169 araD139 relA1 rpsL150 flbB5301 deoC1 tonA21 thi ptsF25 | |
| PCR primers
  MJS-118 | GGG GGG GAA TTC CAT ATG AAC AAA AAA CCC CTA ATC ATT GCC GGC | 5’ primer HAC1267. NdeI site used for cloning is underlined |
| MJS-119 | GGG GGG GTC GAC TCA TCA GCG TTT CAA AAA CGC CTT TAT GAT GG | 3’ primer HAC1267. SalI site used for cloning is underlined |
| MJS-122 | GGG GGG GAA TTC CAT ATG GGT ACA ATT AAA AAA CCC TTA ATC ATT G | 5’ primer HAC1268. NdeI site used for cloning is underlined |
| MJS-123 | GGG GGG GTC GAC TCA TCA TGC AGC CCC CCA TCT AAA ACT AAC | 3’ primer HAC1268. SalI site used for cloning is underlined |
| MJS-154 | GGG GGG GTC GAC TCA TCA CCC CCA AAA AAA TGG TAA AAT TTT ACG CTT TAI TTT TCG | 3’ primer for ΔΔ amino acid at the C-terminal end of HAC1268. SalI site used for cloning is underlined |
| MJS-29 | GGG GGG GTC GAC CAT ATG GAT AAG TTC GCA GAA CAT GAA ATA CCG | 5’ primer for PM1174. NdeI site used for cloning is underlined |
| MJS-31 | GGG GGG GTC GAC CTA CTA AGA CAC GAT GGT CCG TAA TTT TCC | 3’ primer for ΔΔ1 amino acid at the C-terminal end of PM1174. SalI site used for cloning is underlined |
| Plasmids
  pCwori+ | Expression plasmid | Wakarchuk et al. (1994) |
  pCw-MaIE | MalE fusion expression plasmid | Bernatchez et al. (2007) |
  pHAS-02 | HAC1267 full length | This work |
  pHAS-04 | HAC1267 deletion of 4/15 repeats | This work |
  pHAS-07 | HAC1268 deletion of 1/13 repeats | This work |
  pPM03-03 | PM1174 with for ΔΔ1 aa at the C-terminal end | This work |

Cambridge, UK) at 37°C in a microaerobic atmosphere (Thermo Forma, Series II Water Jacketed Incubator; Thermo Fisher Scientific, Waltham, MA). For DNA extraction, *Helicobacter* strains were cultivated in Brain Heart Infusion [Becton, Dickinson and Co. (BD), NJ] containing 10% of fetal bovine serum (Gibco®, Invitrogen, Carlsbad, CA, USA), Skirrow selective supplement (Oxoid) and Vitox supplement (Oxoid) at 37°C in a jar with microaerobic atmosphere. *Escherichia coli* strains were cultivated on the Luria–Bertani medium or 2YT (BIO 101, Carlsbad, CA) supplemented with 150 mg/L of ampicillin when needed. DNA extraction was performed as described previously (Kivisto et al. 2010).
FCHASE(or FEX)-Lactose, FCHASE(or FEX)-LacNAc, FCHASE-Lacto-N-biose and FCHASE-T-Antigen. Other acceptors tested were FCHASE-GaINAc, FCHASE-Gal and FCHASE-GlcNAc. The initial reaction was performed in a final volume of 10 µL in: 50 mM HEPES, pH 7.5, 10 mM MgCl$_2$, 1 mM CMP-5-acetylneuraminic acid (CMP-Neu5Ac), 5 µL of extract diluted in 1 mg/mL of acetylated BSA. Reactions were further optimized with respect to the pH, temperature and metal co-factor. HAC1267 was optimal at pH 8, and HAC1268 was optimal at pH 6.0. We tested 10 mM MgCl$_2$ and 10 mM MnCl$_2$, 5 mM DTT and 2 mM EDTA in various combinations. The reactions were incubated at room temperature and stopped with an equal volume of stop mix (50% acetonitrile, 1% SDS, 10 mM EDTA), then diluted to 1 µM and analyzed by CE as described previously (Wakarchuk and Cunningham 2003).

Molecular modeling of HAC1267 and HAC1268. Initial protein sequence alignments were performed with Clustal X version 2.0 and the resulting alignment was formatted using GeneDoc (Nicholas et al. 1997). Structure models of the two sialyltransferases HAC1267 and HAC1268 were built, based on the sequence alignments and recognized folds, using the fully automated PHYRE2 server (Kelley and Sternberg 2009) hosted at the Imperial College London website, http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index. The two HAC1267 and HAC1268 protein models were successfully generated using the structure of the two sialyltransferases CstI and CstII, as templates. The quality of the generated models was assessed using the QMEAN Server for Model Quality Estimation (http://swissmodel.expasy.org/qmean/cgi/index.cgi). Images were prepared using The PyMOL Molecular Graphics System ver 1.4.1 Schrödinger, LLC.

Supplementary data
Supplementary data for this article are available online at http://glycob.oxfordjournals.org/.

Funding
Drs Warren Wakarchuk and Natalie Strynadka acknowledge funding from the Canadian Institutes for Health Research operating grant MOP-84272. Dr Wakarchuk acknowledges funding from the National Research Council Canada.

Acknowledgements
We thank Dr Evgenii Vinogradov for performing the NMR analysis of the HAC1268 reaction product; Dr Michel Gilbert for critical reading of the manuscript and Dr Mirko Rossi for the gift of the _H. acinonychis_ genomic DNA. Dr Emilie Lameignere is a Cystic Fibrosis Foundation of Canada post-doctoral fellow.

Conflict of interest
None declared.

Abbreviations
BSA, bovine serum albumin; CE, capillary electrophoresis; CMP, cytidine monophosphate; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; FCHASE, 5-(fluorescein-5-carboxamido) hexanoic acid, succinimidyl ester; FEX, 5-fluorescein-EX succinimidyl ester HEPES, N-(2-hydroxyethyl)piperazine-N’-2-ethanesulfonic acid; LOS, lipooligosaccharide; LPS, lipopolysaccharide; Neu5Ac, N-acetylneuraminic acid; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate.

References


