Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure and activity of the *Streptococcus pyogenes* family GH1 6-phospho-β-glucosidase SPy1599

Judith Stepper,[a] Jerome Dabin,[a] Jens M. Eklof,[b] Preeyanuch Thongpoo,[b,c] Prachumporn Kongsaeree,[d] Edward J. Taylor,[a] Johan P. Turkenburg,[a] Harry Brumer[e] and Gideon J. Davies[a]*

[a]Department of Chemistry, University of York, Heslington, York YO10 5DD, England, [b]Division of Glycoscience, School of Biotechnology, Royal Institute of Technology (KTH), AlbaNova University Centre, 106 91 Stockholm, Sweden, [c]Interdisciplinary Graduate Program in Genetic Engineering, Faculty of Graduate School, Kasetsart University, Bangkok 10900, Thailand, [d]Department of Biochemistry, Kasetsart University, Bangkok 10900, Thailand, and [e]Michael Smith Laboratories and Department of Chemistry, University of British Columbia, 2185 East Mall, Vancouver, British Columbia V6T 1Z4, Canada

Correspondence e-mail: gideon.davies@york.ac.uk

Received 31 August 2012
Accepted 28 September 2012

PDB References: SPy1599, 4b3k; 4b3l

The group A streptococcus *Streptococcus pyogenes* is the causative agent of a wide spectrum of invasive infections, including necrotizing fasciitis, scarlet fever and toxic shock syndrome. In the context of its carbohydrate chemistry, it is interesting that *S. pyogenes* (in this work strain M1 GAS SF370) displays a spectrum of oligosaccharide-processing enzymes that are located in close proximity on the genome but that the *in vivo* function of these proteins remains unknown. These proteins include different sugar transporters (SPy1593 and SPy1595), both GH125 α-1,6- and GH38 α-1,3-mannosidases (SPy1603 and SPy1604), a GH84 β-hexosaminidase (SPy1600) and a putative GH2 β-galactosidase (SPy1586), as well as SPy1599, a family GH1 'putative β-glucosidase'. Here, the solution of the three-dimensional structure of SPy1599 in a number of crystal forms complicated by unusual crystallographic twinning is reported. The structure is a classical $(\beta/\alpha)_8$-barrel, consistent with CAZy family GH1 and other members of the GH-A clan. SPy1599 has been annotated in sequence depositions as a β-glucosidase (EC 3.2.1.21), but no such activity could be found; instead, three-dimensional structural overlaps with other enzymes of known function suggested that SPy1599 contains a phosphate-binding pocket in the active site and has possible 6-phospho-β-glycosidase activity. Subsequent kinetic analysis indeed showed that SPy1599 has 6-phospho-β-glucosidase (EC 3.2.1.86) activity. These data suggest that SPy1599 is involved in the intracellular degradation of 6-phosphoglycosides, which are likely to originate from import through one of the organism's many phosphoenolpyruvate phosphotransfer systems (PEP-PTSs).

1. Introduction

Streptococcus pyogenes is an obligate human pathogen and as such has evolved a range of molecular mechanisms to interact with and overcome the human immune system. Much work has been directed towards elucidating the genetic control that these systems are held under, and it has become clear that carbohydrate metabolism plays a pivotal role (Shelburne *et al.*, 2010; Hondorp & McIver, 2007; Shelburne, Davenport *et al.*, 2008). As many as 15% of its genes are thought to be held under carbon catabolite regulation (Kietzman & Caparon, 2010), some of which encode key extracellular virulence factors such as SpeB (a cysteine protease), EndoS (an Ig heavy-chain endoglycosidase) and Spd and Spd3 (DNases; Shelburne, Keith *et al.*, 2008). These secreted proteins cleave the amide bonds of human glycoproteins, the chitobiose core of the asparagine-linked glycan on immunoglobulin G (IgG; Collin & Olsén, 2001) and the phosphodiester bonds of nucleic acids (Suck & Oefner, 1986), respectively. These enzymatic

activities not only debilitate the immune system of the host, but also serve to generate a wide spectrum of sugars and sugar phosphate products. It is not surprising that another response to glucose depletion is the up-regulation of a subset of genes associated with the transport and metabolism of carbohydrates (Kietzman & Caparon, 2010), which reflects this change to a new carbon source.

S. pyogenes strain M1 GAS SF370 open reading frame SPy1599 is part of a cluster of genes involved in carbohydrate metabolism. This gene cluster includes genes encoding SPy1600 (a β-*N*-acetylglucosaminidase; Sheldon *et al.*, 2006), SPy1602 (a possible transcription regulator), SPy1603 (an α-mannosidase), SPy1604 (an α-mannosidase; Suits *et al.*, 2010) and SPy1593 and SPy1595 (both putative sugar-binding transport proteins). The above gene products all hint at a role in the degradation and utilization of N-glycans, but the activity and role of SPy1599 is unclear. At the sequence level, SPy1599 is classified into CAZy family GH1 (Cantarel *et al.*, 2009), the characterized members of which are predominantly variants of β-glucosidases and β-galactosidases (see http://www.cazy.org/GH1.html).

Here, we present the three-dimensional structure of SPy1599 in a variety of crystal forms; the structure solution was complicated by crystallographic twinning. We show that SPy1599 has 6-phospho-β-glucosidase activity, consistent with three-dimensional structural similarity searches and the convergent evolution of an active-centre (likely phosphate-coordinating) lysine.

2. Materials and methods

2.1. Gene cloning, expression and protein purification

The *S. pyogenes* putative β-glycosidase (SPy1599) gene (Gene ID 901846) was amplified by the polymerase chain reaction (PCR). The primers used were 5′-**GACGACGACAAGAT**GTTAGCATTTCCAAAAGAGTT-3′ (forward primer) and 5′-**GAGGAGAAGCCCGGT**TCAATCACTGAGGTGAGAAG-3′ (reverse primer). The bold sections of the primers are ligation-independent cloning (LIC) sites, which were used to clone the PCR product into pET30 Ek/LIC (Novagen). This yielded an SPy1599 construct containing a 6×His sequence at the N-terminus, hereafter called pET30-SPy1599-His_6.

Competent *Escherichia coli* BL21 (DE3) cells were transformed by heat shock with pET30-SPy1599-His_6 and selected on LB–agarose plates containing 30 µg ml^{-1} kanamycin. Transformed cells were grown in auto-induction medium at 310 K to an OD_{600} of about 3 and were then allowed to grow overnight at 303 K. The cells were harvested by centrifugation (5000 rev min^{-1}, 20 min), resuspended in 50 m*M* NaH_2PO_4 pH 7.0, 300 m*M* NaCl, 10 m*M* imidazole and lysed by sonication. A mixture of proteinase inhibitors (Complete Mini) was added and the insoluble fraction of the lysate was removed by centrifugation at 18 000 rev min^{-1} for 30 min. The supernatant was loaded onto an Ni^{2+} agarose column (HisTrap FF, Amersham Biosciences) and SPy1599-His_6 was eluted using an imidazole gradient, resulting in 90% pure protein as estimated by SDS–PAGE. SPy1599-His_6-containing fractions were pooled and concentrated using a 30 kDa cutoff polyethersulfone membrane concentration device (Vivascience). The concentrated protein sample was further purified by gel filtration using a HiLoad 16/60 Superdex 200 prep-grade column eluted with 50 m*M* HEPES pH 7.0, 150 m*M* NaCl. The purification yielded about 60 mg Spy1599. SDS gels showed that SPy1599 was subject to degradation during both storage and crystallization experiments.

Table 1
X-ray data and structure-refinement statistics.

	*C*2	$P2_1$, twinned†	$P2_1$, not twinned
Data collection			
Unit-cell parameters			
a (Å)	109.2	106.4	107.8
b (Å)	186.0	190.2	198.0
c (Å)	101.2	106.2	107.9
$\alpha = \gamma$ (°)	90.0	90.0	90.0
β (°)	108.3	118.9	118.5
Molecules in asymmetric unit	3	6	6
Resolution (Å)	2.61 (2.68–2.61)	2.60 (2.67–2.60)	2.51 (2.58–2.51)
R_{merge}	0.16 (1.10)	0.09 (0.62)	0.063 (0.46)
Mean $I/\sigma(I)$‡	8.4 (2.1)	12.4 (1.9)	10.7 (2.3)
Completeness (%)	99.0 (99.3)	94 (83)	99.9 (100.0)
Multiplicity	3.3 (3.2)	3.5 (3.2)	3.7 (3.8)
Refinement			
R_{cryst}/R_{free} (%)	Not refined	22.3/28.6	20.5/26.1
R.m.s.d. bonds (Å)		0.012	0.014
R.m.s.d. angles (°)		1.6	1.7
Mean *B*, protein (Å^2)		39.4	46.3
PDB code	n/a	4b3k	4b3l

† Pseudo-merohedrally twinned with twin operator $l, -k, h$ and refined twin fraction of 0.33 (from *REFMAC*). ‡ As defined by *SCALA* (Evans, 2006).

2.2. X-ray crystallography and structure solution

SPy1599 was crystallized by vapour diffusion using the hanging-drop method from two conditions: 0.1 *M* bis-Tris propane pH 7.5, 0.2 *M* sodium fluoride or sodium bromide, 20% polyethylene glycol 3350. Crystals were obtained at 292 K using equal volumes of protein solution (20 mg ml^{-1}) and mother liquor. These conditions yielded several related crystal forms ($P2_1$, *C*2 and $C222_1$, with several different unit-cell parameter changes for each form). Two crystal forms, $P2_1$ (NaF) and *C*2 (NaBr), of apo SPy1599 were obtained and were ultimately used for structure solution (below). Crystals were cryoprotected in a solution containing the mother liquor with 22%(*v*/*v*) ethylene glycol and were harvested into rayon-fibre loops prior to flash-cooling in liquid nitrogen. X-ray data were collected at the Diamond Light Source and were processed using the *xia*2 (Winter, 2010) implementations of *XDS* (Kabsch, 1993) or *MOSFLM* (Leslie, 2006).

Structure solution of SPy1599 proved to be extremely challenging as the initial $P2_1$ space-group data sets, which were likely to have 6–8 molecules in the asymmetric unit, displayed a high degree of pseudo-merohedral twinning (twinning statistics are discussed further in §3.2 below). The structure was eventually solved by searching for an untwinned data set,

in this case a *C*2 crystal form with unit-cell parameters $a = 109.2$, $b = 186.0$, $c = 101.2$ Å, $\beta = 108.3^\circ$. Although the data were of medium resolution and poor quality (Table 1; $R_{\rm merge}$ of 1.10 in the outer bin), these data allowed successful structure solution by molecular replacement. The structure was solved by molecular replacement using the *CCP*4 (Winn *et al.*, 2011) implementation of *MOLREP* (Vagin & Teplyakov, 2010) and with a single molecule of the *Thermatoga maritima Tm*GH1 β-glucosidase as the search model (PDB entry 1od0; Zechel *et al.*, 2003). The initial solution with three molecules in the asymmetric unit had starting $R_{\rm cryst}$ and $R_{\rm free}$ values in excess of 51% at 2.6 Å resolution, which only dropped to 48% and 50%, respectively, during refinement with *REFMAC* (Murshudov *et al.*, 2011). The model was therefore improved, and the correct sequence was modelled, using *Buccaneer* (Cowtan, 2006), which initially built around 1000 residues (out of a total of approximately 1300) but with many main-chain breaks. Phase improvement using *Parrot* (Cowtan, 2010), featuring non-crystallographic symmetry averaging of the three molecules, allowed a far more robust building with *Buccaneer*, with all three molecules built with a connectivity of 460 amino acids (of 463) in each. This structure was refined using *REFMAC* to yield a protein-only model with $R_{\rm cryst}$ and $R_{\rm free}$ values of 29% and 31%, respectively. A single molecule from this interim model was used to solve the (twinned) $P2_1$ form, but given the very poor data quality (Table 1) the *C*2 model was not used subsequently.

The first $P2_1$ crystal form (unit-cell parameters $a = 106.4$, $b = 190.2$, $c = 106.2$ Å, $\beta = 118.9^\circ$) provided a much better quality data set (Table 1). Initial structure solution of this form was hindered by the combination of many molecules in the asymmetric unit and the presence of pseudo-merohedral twinning. Using the *C*2 model, however, structure solution using *MOLREP* proved facile, yielding a solution with six molecules in the asymmetric unit. Model building using *Coot* (Emsley & Cowtan, 2004) and refinement using *REFMAC* featuring detwinning (discussed in §3) yielded a final model with $R_{\rm cryst}$ and $R_{\rm free}$ values of 22% and 28%, respectively.

During the completion of the above refinement, a second $P2_1$ crystal form was obtained that was closely related to that described above. This form had unit-cell parameters $a = 107.8$, $b = 198.0$, $c = 107.9$ Å, $\beta = 118.5^\circ$. Data were collected to 2.5 Å resolution; the structure was solved using *MOLREP* and refinement using *REFMAC* proceeded smoothly. These data displayed no twinning. Final model data statistics are given in Table 1.

2.3. Substrates

p-Nitrophenyl-β-D-glucopyranoside (*p*NP-Glc), *p*-nitrophenyl-β-D-galactopyranoside (*p*NP-Gal) and *p*-nitrophenyl-β-D-xylopyranoside (*p*NP-Xyl) were purchased from Sigma–Aldrich, St Louis, Missouri, USA. *p*-Nitrophenyl-*N*-acetyl-β-D-glucosamine (pNP-GlcNAc) and *p*-nitrophenyl-*N*-acetyl-β-D-galactosamine (pNP-GalNAc) were purchased from Carbosynth, Berkshire, England. The phosphorylated substrate *p*-nitrophenyl 6-phospho-β-D-galactopyranoside (pNP-Gal6P, cyclohexylamine salt) was purchased from Gold Biotechnology, St Louis, Missouri, USA. *p*-Nitrophenyl 6-phospho-β-D-glucopyranoside (pNP-Glc6P) was obtained from Professor Stephen G. Withers, Vancouver, Canada.

2.4. Enzyme kinetics and pH profile of catalysis

The protein content of the SPy1599 sample was determined from the A_{280}. For all calculations, the total enzyme concentration $[E]_{\rm t}$ was assumed to be equivalent to the total protein concentration.

The pH optimum of SPy1599 towards pNP-Glc6P (final concentration 1 m*M*) was determined in 100 m*M* sodium citrate pH 3.0–6.4 and 100 m*M* sodium 4-morpholinepropanesulfonic acid (MOPS) pH 6.4–8.0. The reactions (total volume 100 µl), which contained 0.73 µ*M* SPy1599, were incubated at 310 K for 13 min. The *p*-nitrophenol released was determined from the A_{400} (Cary 50, Varian) using an extinction coefficient of $\varepsilon = 18\,500\ M^{-1}\,{\rm cm}^{-1}$ after stopping the assays by the addition of 400 µl 0.2 *M* Na_2CO_3. All measurements were performed in duplicate.

The kinetic parameters $k_{\rm cat}$ and $K_{\rm m}$ of SPy1599 were determined for pNP-Glc6P from the rate of *p*-nitrophenol production in the concentration range 0.1–5 m*M*. Kinetic measurements were performed at the pH optimum in 100 m*M* sodium citrate buffer pH 6 at 310 K with 0.36 µ*M* SPy1599 (total reaction volume 100 µl). After 13 min incubation, the reactions were stopped by adding 400 µl 0.2 *M* Na_2CO_3 and the amount of *p*-nitrophenol was determined from the A_{400} (Cary 50, Varian). All reactions were performed in duplicate. The kinetic parameters were determined by nonlinear curve fitting of the standard Michaelis–Menten equation using the software program *Origin* v.8. The relative specific activities of SPy1599 were determined against 2 m*M* pNP-Glc, pNP-Gal, pNP-Xyl, pNP-GlcNAc and pNP-GalNAc using a similar methodology with an assay time of 32 min and 9.6 µ*M* SPy1599. Activity towards 1.25 m*M* pNP-Gal6P was assayed in a similar way overnight at a protein concentration of 0.36 µ*M* SPy1599.

To test whether cyclohexylammonium (or its conjugate base), which is present as a counterion in the pNP-Glc6P substrate, could act as an inhibitor of SPy1599, 0.36 µ*M* of the protein in an assay volume of 100 µl containing 1.25 m*M* pNP-Glc6P in 100 m*M* sodium citrate pH 6 was incubated at 310 K for 17 min with and without added cyclohexylamine (2.5 m*M*). The release of *p*-nitrophenol was then monitored from the A_{400} after the addition of 400 µl 0.2 *M* Na_2CO_3.

3. Results

3.1. Gene organization

The genomic location of the SPy1599 locus is associated with a number of other genes related to carbohydrate metabolism. The genes within this cluster can be roughly divided into those which encode glycoside hydrolases (SPy1599, SPy1600, SPy1603 and SPy1604), regulatory proteins (SPy1596 and SPy1602) and transporter proteins (SPy1593

and SPy1595). All of the genes within the cluster are devoid of sequences which encode signal peptides; as such, we tentatively suggest that the proteins that they encode are intracellular.

Amino-acid sequence similarities show that Spy1600 is a CAZy GH84 enzyme (http://www.cazy.org). Spy1600 is a β-*N*-acetylglucosaminidase (Sheldon *et al.*, 2006); its *in vivo* substrate is unknown, but it is able to cleave the β-linked GlcNAc linked to serine and threonine residues as well as GlcNAc-containing aryl glycosides and oligosaccharides *in vitro*. SPy1604 is an α-mannosidase belonging to GH38; it shows specificity for α-1,3- and α-1,6-mannosyl linkages (Suits *et al.*, 2010), which are frequently associated with N-glycans. SPy1603 shows 62% identity (*BLAST*; http://blast.ncbi.nlm.nih.gov) to a GH125 α-mannosidase from *S. pneumoniae* (SpGH125; Gregg *et al.*, 2011). The SpGH125 protein shows exo-α-1,6-mannosidase activity, which is again consistent with specificity for N-linked glycans. All of these hydrolytic activities are consistent with the catabolic downstream processing of imported complex carbohydrates, especially N-glycans.

SPy1596 shows 37% identity (*BLAST*) to a ROK (repressor, open reading frame, kinase) protein in complex with sucrose from *S. pneumoniae* Tigr4. ROK proteins act as responsive transcriptional repressors and regulate carbohydrate metabolism by acting as sugar sensors, with saccharide phosphorylation providing a signal for shifts in metabolism (Conejo *et al.*, 2010). SPy1602 shows identity to the GntR family of transcriptional regulators. These proteins have variable DNA-binding capacities when in complex with monosaccharides such as lactose, ribose, fructose and xylose that regulate gene expression. SPy1593 and SPy1595 encode the P-binding cassette and transmembrane subunit of an ATP-binding cassette (ABC) transporter. These are typically oligosaccharide transporters comprising two integral membrane permease proteins, two peripheral ATP-hydrolysing membrane proteins and an extracytoplasmic solute-binding receptor (Ajdić & Pham, 2007).

Taken together, the products of this gene cluster show a high degree of functional synergy and potentially represent a suite of molecular tools for the downstream processing of imported complex glycans. We therefore sought to determine whether SPy1599 would fit into such a scheme. Unfortunately, a screen of classical β-aryl glycoside substrates failed to reveal any significant activity towards pNP-Xyl, pNP-Glc, pNP-GlcNAc, pNP-Gal or pNP-GalNAc. These data clearly show that SPy1599 is a not a β-glucosidase as annotated (for example, in deposited sequences such as UniProtKB/TrEMBL entry Q99YP9), nor is it active towards other closely related glycosides. This lack of obvious activity promoted us to pursue its structure determination.

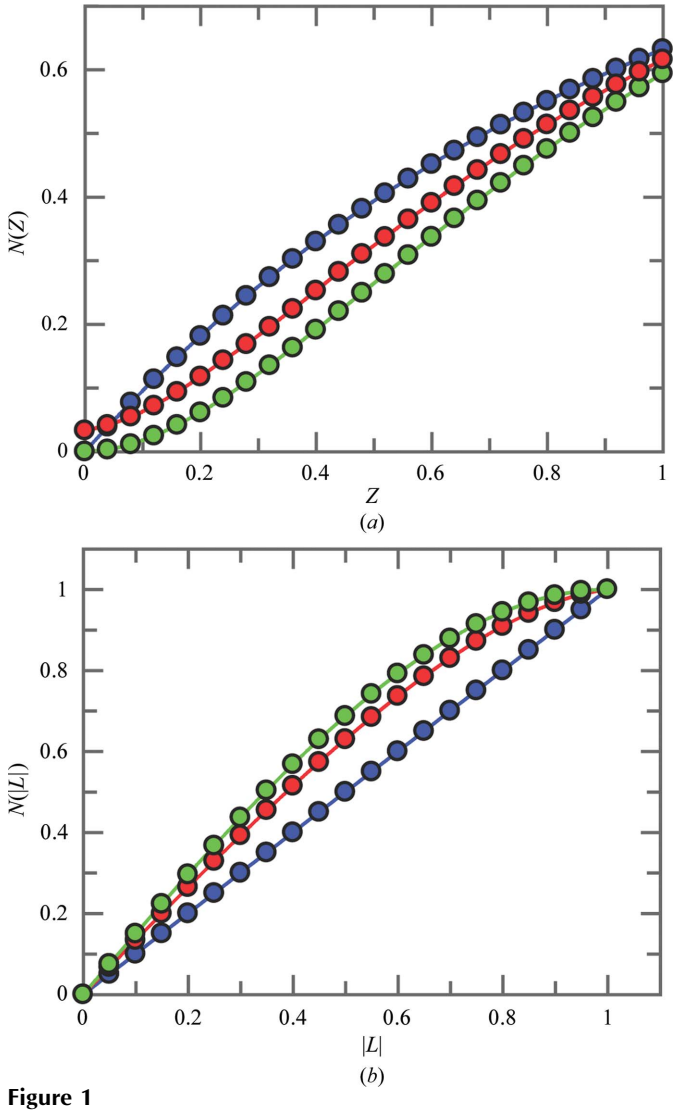

Figure 1
Analysis of crystal twinning for the pseudo-merohedrally twinned $P2_1$ form. (*a*) Cumulative intensity distribution for acentric data only; (*b*) *L*-test (Padilla & Yeates, 2003). In both graphs the theoretical untwinned data are shown as blue circles, the theoretical untwinned data as green circles and the observed data as red circles.

3.2. Structure solution and three-dimensional structure of SPy1599

At the time of structure solution, the ORF SPy1599 displayed high sequence similarity to other proteins annotated as putative β-glucosidases but far less similarity to proteins of known three-dimensional structure; the closest model was the *T. maritima* β-glucosidase *Tm*GH1, with just 33% sequence identity. Structure solution with a low-sequence-similarity model was hindered by the combination of both a large asymmetric unit volume (suggesting 6–10 molecules in the $P2_1$ asymmetric unit) and the presence of very clear pseudo-merohedral twinning in the $P2_1$ data, as reflected in the cumulative intensity distribution, *L*-test and moment analyses of the *CCP*4 *TRUNCATE*/*POINTLESS* programs (Fig. 1). The data can also be reduced in space group $C222_1$, albeit with slightly worse statistics, presumably reflecting the imperfect (twin fraction 0.33) nature of the twin (Table 1). In $C222_1$ the overall $R_{\rm merge}$ is 0.15 compared with 0.088 in $P2_1$; crucially, that in the inner resolution bin jumps from 0.026 in $P2_1$ to 0.083 in $C222_1$. At this point, putative solutions, with poor electron-density maps, could not be pursued with any confidence.

The breakthrough in the structure solution came from the systematic analysis of a large number of SPy1599 data sets. In the end, a single untwinned data set in space group *C*2 was found with an asymmetric unit volume implying the presence of 3–4 molecules in the asymmetric unit. Despite the extremely poor-quality data, the *C*2 structure could be solved using *MOLREP* with the *Tm*GH1 coordinates as a search model, generating an unambiguous three-molecule solution. The initial structure proved difficult to refine, but semi-automatic model building was possible using *Buccaneer* with the phases first improved using threefold averaging with the *Parrot* program. This automatically built structure was refined using *REFMAC* to yield an interim model with $R_{\rm cryst}$ and $R_{\rm free}$ values of 0.27 and 0.30, respectively; this structure was itself not refined further, but was instead used to solve the structure of the twinned $P2_1$ form (Table 1). Subsequently, a further data set was obtained with six possible molecules in the asymmetric unit. Its untwinned $P2_1$ structure was solved to 2.51 Å resolution using the unrefined *C*2 structure (Table 1, Fig. 2).

Glycoside hydrolases are organized into clans and families according to similarities in tertiary structure and conservation of the catalytic residues and catalytic mechanism (Henrissat & Bairoch, 1996; Cantarel *et al.*, 2009). The structure of SPy1599 (Fig. 2*a*) presents the typical $(\beta/\alpha)_8$-barrel fold of the GH1 family belonging to clan GH-A (Henrissat *et al.*, 1995). It is composed of 18 α-helices and 12 β-sheets. The catalytic acid/base Glu165 is located in the NEP motif at the end of β4 and the catalytic nucleophile Glu366 is located in the ENG motif at the end of β7. The residues surrounding the active site of SPy1599 comprise the typical residues for GH1 enzymes: Arg74, His119, Asn164, Asn297, Tyr299 and Trp416. Five *cis*-peptide bonds were found at Gly11–Gly12, Tyr180–Pro181, Ala287–Glu288, Trp416–Thr417 and Ser423–Trp424 of SPy1599. The occurrence of such bonds is characteristic of glycoside hydrolase family 1 (as discussed, for example, in Seshadri *et al.*, 2009).

At the time of writing, structure-similarity searches using *PDBeFold* (*SSM* at the EBI) shows that SPy1599 is, as expected, closely related to other members of CAZy family GH1. The closest matches (ranked by the default option of *SSM Q* score) are for a *Halothermothrix orenii* GH1 (PDB entry 3ta9; r.m.s.d. of 1.5 Å over 432 matched C^α atoms with 33% sequence identity; Kori *et al.*, 2011), a *Streptomyces* β-glucosidase (r.m.s.d. of 1.4 Å over 423 matched C^α atoms with 31% sequence identity; PDB entry 1gon; A. Guasch, M. Vallmitjana, R. Perez, E. Querol, J. A. Perez-Pons & M. Coll, unpublished work), the *T. maritima* β-glucosidase *Tm*GH1 (r.m.s.d. of 1.5 Å over 426 matched C^α atoms with 33% sequence identity; PDB entry 2wc4; Aguilar-Moncayo *et al.*, 2009) and a *Paenobacillus polymyxa* β-glucosidase (r.m.s.d. of 1.5 Å over 429 matched C^α atoms with 31% sequence identity; PDB entry 1uyg; Wright *et al.*, 2004). The structural similarities are all very close in score and thus are subject to the ranking algorithm; for example, when the *SSM P*-score was used the closest structure was the 6-phospho-β-glucosidase from *Lactobacillus lactis* (r.m.s.d. of 1.4 Å over 420 matched C^α atoms with 30% sequence identity; PDB entry 1pbg; Wiesmann *et al.*, 1995).

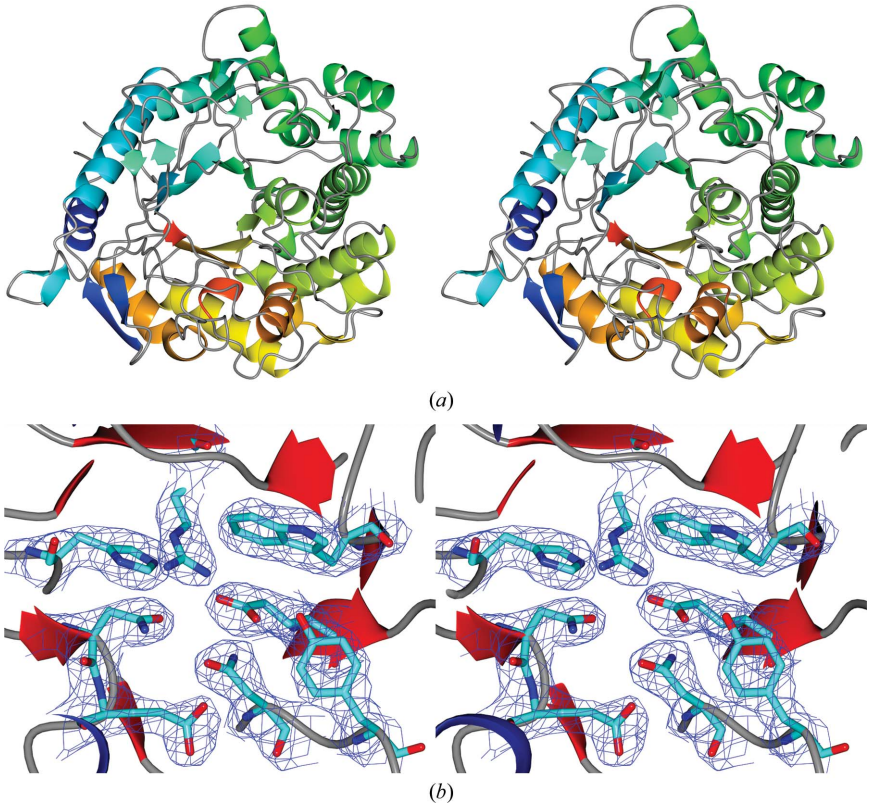

Figure 2
Three-dimensional structure of SPy1599. (*a*) Three-dimensional protein cartoon colour-ramped from the N-terminus to the C-terminus. (*b*) Sample electron density for the *Coot*-averaged $2F_{\rm o} - F_{\rm c}$ electron-density map in the vicinity of the active centre for the $P2_1$ untwinned data. This figure was drawn with *CCP*4*mg* (McNicholas *et al.*, 2011).

3.3. Active centre and potential specificity

Our initial kinetic analysis failed to reveal the expected β-glucosidase activity whilst the structural similarity search revealed a large number of equivalent structural homologues, many of which were proteins of unknown (or inferred) function. Clues to the activity of SPy1599 came from three-dimensional structural superimposition with GH1 enzymes of known function identified as above. Especially striking were the overlays in the −1 subsite of the enzyme, where SPy1599 clearly had an additional pocket around the substrate O6 position compared with classical β-glucosidases such as *Tm*GH1. Furthermore, comparison with 6-phospho-β-glucosidase from *L. lactis* (PDB entry 1pbg) revealed that the 6-phospho group of this structure lay in the pocket observed in SPy1599 that was absent from classical β-glucosidases,

Table 2
Kinetic properties of SPy1599.

Substrate	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1} mM^{-1})	$v_0/[E]_t$† (min^{-1})	Relative $v_0/[E]_t$† (%)
pNP-Glc6P	0.73 ± 0.018	14.57 ± 0.13	20	11 ± 0.04	100
pNP-Gal	—	—	—	0.21± 0.01	2
pNP-Glc	—	—	—	0.0069 ± 0.00009	<0.1
pNP-Xyl	—	—	—	0.008 ± 0.0012	<0.1
pNP-Gal6P	—	—	—	<0.001	—
pNP-GalNAc	—	—	—	<0.001	—
pNP-GlcNAc	—	—	—	<0.001	—

† Determined at [S] = 2 mM for pNP-Glc, pNP-Gal, pNP-Xyl, pNP-GlcNAc and pNP-GalNAc.

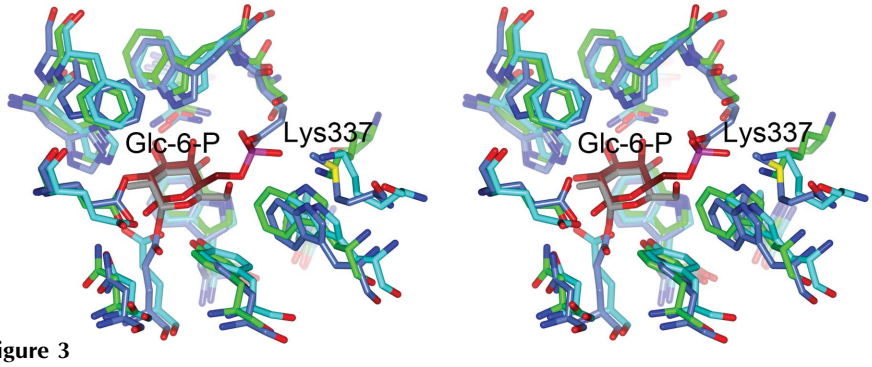

Figure 3
Comparison of the active site of SPy1599 with those of β-glucosidase from *T. maritima* (*Tm*GH1) and 6-phospho-β-glucosidase from *L. lactis*. Stereoview of the active-site residues of SPy1599 (cyan) superpositioned with *Tm*GH1 (PDB entry 1oin; light blue; Zechel *et al.*, 2003) and 6-phospho-β-glucosidase from *L. lactis* (PDB entry 4pbg; green with ligand in dark red; Wiesmann *et al.*, 1997). This figure was prepared using *CCP4mg* (McNicholas *et al.*, 2011).

suggesting that SPy1599 may function as a 6-phospho-β-glycosidase.

Further evidence that SPy1599 may use 6-phospho substrates comes from the positioning of a lysine side chain in the active centre of SPy1599 and 6-phospho-β-glucosidase from *L. lactis* (PDB entry 1pbg). The characterized 6-phospho-β-glucosidase of known structure features a lysine side chain which interacts directly with the phosphate moiety of the substrate (Fig. 3). In SPy1599 there is not a lysine at the same position in the primary structure; instead, a different lysine, Lys337, from elsewhere in the three-dimensional structure is positioned with its terminal amino group in the equivalent position. This strongly suggests convergent evolution to produce a phosphate-binding charge–charge interaction. Furthermore, if one performs sequence searches only on proteins of published and characterized function (available at http://www.cazy.org/GH1_characterized.html), one sees that a lysine or arginine side chain is conserved in all of the closest homologues of SPy1599 where they are known to be 6-phospho-β-glycosidases. In light of this structural revelation, the activity of SPy1599 was reassessed towards both pNP 6-phospho-β-D-glucoside (pNP-Glc6P) and pNP 6-phospho-β-D-galactoside (pNP-Gal6P).

3.4. SPy1599 is a 6-phospho-β-glucosidase

SPy1599 was active towards pNP 6-phospho-β-D-glucoside following classical Michaelis–Menten kinetics (Table 2, Fig. 4*a*). Using this substrate, the pH profile of the enzyme had an optimum of ∼6 and acidic and basic limbs having pK_a values of 5.6 ± 0.2 and 7.0 ± 0.2, respectively (Fig. 4*b*). Surprisingly, SPy1599 did not show any hydrolytic activity towards pNP-Gal6P. pNP-Gal6P is supplied as a cyclohexylamine salt containing two equivalents of the base. To test whether the amine might inhibit SPy1599, an additional amount was added in an assay with pNP-Glc6P. An additional two equivalents of base (2.5 mM) relative to substrate (1.25 mM) did not show any inhibitory effect on the specific activity of SPy1599, thereby ruling out amine/ammonium inhibition as a cause of the lack of hydrolytic activity towards pNP-Gal6P. After careful re-examination of the activity of SPy1599 on pNP glycosides, some hydrolytic activity could however be observed on pNP-Gal (2% of the specific activity on pNP-Glc6P at 2 mM; Table 2).

The data clearly show that SPy1599 has 6-phospho-β-glucosidase activity, as predicted from the three-dimensional structure. This activity corrects the putative homology-based (automated) assignment of function, but poses conundrums when it comes to assigning a biological function for the SPy1599 gene product.

4. Discussion

The location of several glycosidase genes, as well as sugar transporters, in close proximity on the *S. pyogenes* genome strongly hints at a role in oligosaccharide or polysaccharide metabolism. Furthermore, the presence of both α-1,6- and α-1,3-mannosidases as well as a β-*N*-acetylglucosaminase implies a role in N-glycan degradation and utilization. Such a role would be consistent with the role of EndoS in removing N-glycans as part of the immune-evasion strategy of the organism (Collin & Olsén, 2001). Arguing against this is the absence of signal peptides from all of these proteins, apart from EndoS, implying cytoplasmic location (Sheldon *et al.*, 2006; Suits *et al.*, 2010). Conversely, there are also recent data showing that other *S. pyogenes* proteins with no discernible secretion sequence are indeed known to exist in an extracellular location (Jin *et al.*, 2011; Pancholi & Fischetti, 1998), so one must acknowledge the possibility of a previously uncharacterized protein-export system in *S. pyogenes* (Pancholi & Fischetti, 1998). Whilst an uncharacterized protein-export machinery may play a role, there is however little direct evidence to support such a role for these enzymes in N-glycan degradation. In the case of SPy1599, based on the commercially available aryl glycoside substrates used, we have

shown that SPy1599 is primarily a 6-phospho-β-glucosidase, yet 6-phospho-β-D-glucosides are not part of the N-glycan repertoire of any organism. The most likely origin of a 6-phospho-β-D-glucoside is through the phosphotransferase systems (PTSs), which simultaneously phosphorylate and import short oligosaccharides.

The PTS is a multi-component system common to many bacterial species including *S. pyogenes*. Sugars which are external to the cell are transferred across a cell membrane by means of a specific phosphorylating permease. This process is driven by the addition of phosphate to the sugars to create a transport gradient. The sugars in the cytosol are processed by a suite of intracellular enzymes during assimilation. Specific components of the *S. pyogenes* PTSs which have been shown to be held under carbon catabolite regulation are involved in the uptake and utilization of galactose, mannose, lactose and fructose (Shelburne, Keith *et al.*, 2008), including β-glucoside-specific transporters such as SPy0572 (Kietzman & Caparon, 2010), which could potentially provide substrates for SPy1599.

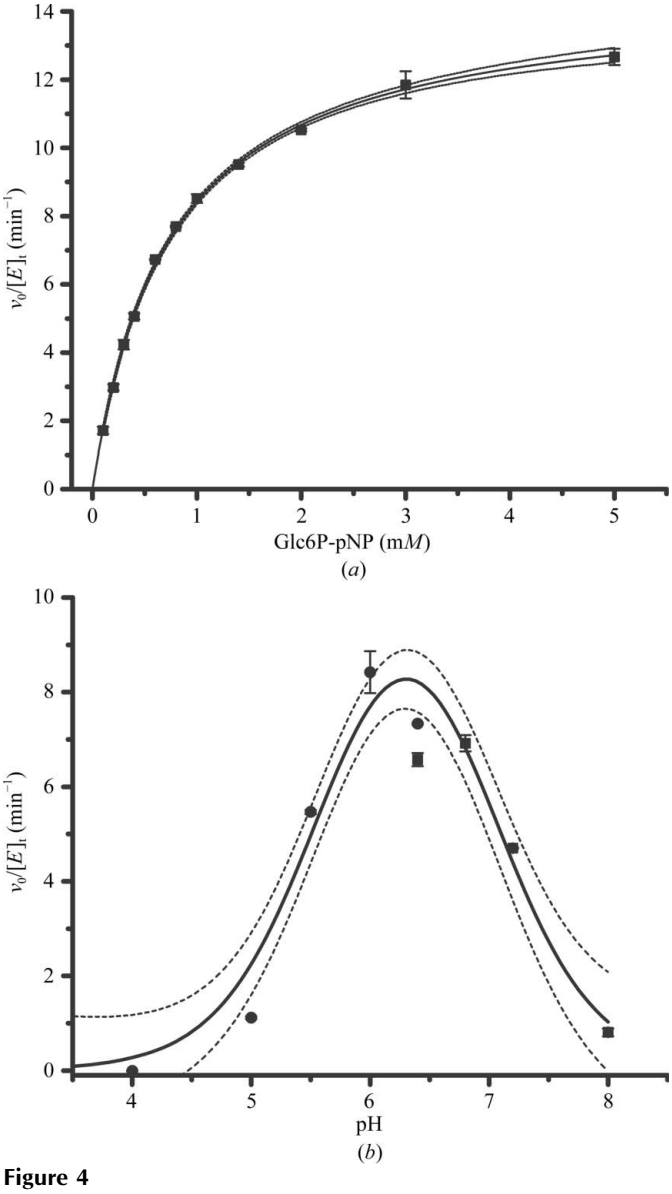

Figure 4
Initial rate kinetics of SPy1599 on pNP-Glc6P. (*a*) A Michaelis–Menten fit of the initial rate kinetics on pNP-Glc6P. Error bars are two standard deviations; the 95% confidence interval of the fitted curve is shown as dotted lines. (*b*) The pH–rate profile of SPy1599 on pNP-Glc6P, showing the effect of pH on SPy1599-catalysed hydrolysis of 1 mM pNP-Glc6P between pH 3.0 and 8.0. Filled black circles, sodium citrate buffer; filled black squares, MOPS buffer. Error bars are two standard deviations, representing a 95% confidence interval. The 95% confidence interval of the fitted curve is shown as dotted lines.

Indeed, the KEGG pathways database (http://www.genome.jp/kegg/pathway.html) has identified an even wider range of PTS transporter proteins within the *S. pyogenes* M1 GAS SF370 genome. These are likely to enable the uptake of sucrose (SPy1815), trehalose (SPy2097 and SPy1986), lactose (SPy1917 and SPy1918), cellobiose (SPy2050, SPy1323 and SPy1324), mannose (SPy1059, SPy1060 and SPy1057), *N*-acetylgalactosamine (SPy0630, SPy0629, SPy0634 and SPy0631), galactitol (SPy1709, SPy1711 and SPy1710) and fructose (SPy0855). Together, these predictions suggest that the PTS is the general mechanism for the uptake of short oligosaccharides in M1 GAS SF370. Furthermore, *S. pyogenes* has no CAZy family GH4 6-phosphoglycosidases that are often associated with phosphoenolpyruvate phosphotransfer systems (PEP-PTSs) in other organisms, so an alternative enzyme or enzymes must exist in this organism.

There are few clues as to the real function of SPy1599; both SPy1599 and the hexosaminidase SPy1600 are downregulated when *S. pyogenes* becomes adhered to pharyngeal cells (Ryan *et al.*, 2007). As these are, to our knowledge, the only two genes from the cluster known to exhibit co-transcriptional regulation, it is likely that their function is connected and perhaps discrete from the rest of the cluster. Interestingly, SPy1599 and some locus-associated genes SPy1593, SPy1596, SPy1600, SPy1602 and SPy1604 are all repressed in cells which are cultured in the presence of glucose (Kietzman & Caparon, 2010) and all bar SPy1604 are regulated by catabolite control protein A (CcpA; Kietzman & Caparon, 2010). This points to SPy1599 (and co-expressed genes) having a role related to the processing of carbohydrates acquired at a time of glucose depletion. In *S. pyogenes*, CcpA also regulates the expression of a number of extracellular virulence factors (Hondorp & McIver, 2007) active in carbohydrate foraging (Shelburne, Davenport *et al.*, 2008). In addition to this, the CcpA regulon protein also controls gene expression of the PTS (Fujita, 2009) and ABC transporters, which reflect the change to a new carbon source.

Thus, the true role of the closely associated genes remains unclear. Many genes that would assist in N-glycan degradation are present, but not all. *S. pyogenes* appears to possess neither a fucosidase nor a sialidase. Similarly, one might envisage roles in mucin degradation either for the purpose of biological nutrition or overcoming mucosal barriers (Ryan *et al.*, 2001), but again many of the genes that one would consider to be necessary are absent. It is possible that *S. pyogenes* only has certain members of these pathways, while members of the upper respiratory tract microbial community contribute the remaining enzymes, collectively leading to complete pathways. The exact role of the SPy1599 gene product may require

extensive transcriptomic approaches for unambiguous dissection.

This work was funded by a grant from the Biotechnology and Biological Sciences Research Council (BBSRC; Grant BB/G016127/1). EJT is a Royal Society University Research Fellow. HB acknowledges financial support from The Swedish Research Council Formas *via* 'CarboMat – The KTH Advanced Carbohydrate Materials Consortium' and start-up funding from the University of British Columbia. PT and PTK were supported by the Commission on Higher Education and the National Research Council of Thailand. Andrey Lebedev is thanked for useful discussions with JS concerning twinning. The staff of the Diamond Light Source are thanked for the provision of data-collection facilities.

References

Aguilar-Moncayo, M., Gloster, T. M., Turkenburg, J. P., García-Moreno, M. I., Ortiz Mellet, C., Davies, G. J. & García Fernández, J. M. (2009). *Org. Biomol. Chem.* **7**, 2738–2749.
Ajdić, D. & Pham, V. T. (2007). *J. Bacteriol.* **189**, 5049–5059.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Collin, M. & Olsén, A. (2001). *EMBO J*, **20**, 3046–3055.
Conejo, M. S., Thompson, S. M. & Miller, B. G. (2010). *J. Mol. Evol.* **70**, 545–556.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Cowtan, K. (2010). *Acta Cryst.* D**66**, 470–478.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Fujita, Y. (2009). *Biosci. Biotechnol. Biochem.* **73**, 245–259.
Gregg, K. J., Zandberg, W. F., Hehemann, J. H., Whitworth, G. E., Deng, L., Vocadlo, D. J. & Boraston, A. B. (2011). *J. Biol. Chem.* **286**, 15586–15596.
Henrissat, B. & Bairoch, A. (1996). *Biochem. J.* **316**, 695–696.
Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. & Davies, G. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 7090–7094.
Hondorp, E. R. & McIver, K. S. (2007). *Mol. Microbiol.* **66**, 1056–1065.
Jin, H., Agarwal, S. & Pancholi, V. (2011). *MBio*, **2**, e00068-11.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Kietzman, C. C. & Caparon, M. G. (2010). *Infect. Immun.* **79**, 812–821.
Kori, L. D., Hofmann, A. & Patel, B. K. C. (2011). *Acta Cryst.* F**67**, 111–113.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.
Pancholi, V. & Fischetti, V. A. (1998). *J. Biol. Chem.* **273**, 14503–14515.
Ryan, P. A., Kirk, B. W., Euler, C. W., Schuch, R. & Fischetti, V. A. (2007). *PLoS Comput. Biol.* **3**, e132.
Ryan, P. A., Pancholi, V. & Fischetti, V. A. (2001). *Infect. Immun.* **69**, 7402–7412.
Seshadri, S., Akiyama, T., Opassiri, R., Kuaprasert, B. & Cairns, J. K. (2009). *Plant Physiol.* **151**, 47–58.
Shelburne, S. A., Davenport, M. T., Keith, D. B. & Musser, J. M. (2008). *Trends Microbiol.* **16**, 318–325.
Shelburne, S. A., Keith, D., Horstmann, N., Sumby, P., Davenport, M. T., Graviss, E. A., Brennan, R. G. & Musser, J. M. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 1698–1703.
Shelburne, S. A., Olsen, R. J., Suber, B., Sahasrabhojane, P., Sumby, P., Brennan, R. G. & Musser, J. M. (2010). *PLoS Pathog.* **6**, e1000817.
Sheldon, W. L., Macauley, M. S., Taylor, E. J., Robinson, C. E., Charnock, S. J., Davies, G. J., Vocadlo, D. J. & Black, G. W. (2006). *Biochem. J.* **399**, 241–247.
Suck, D. & Oefner, C. (1986). *Nature (London)*, **321**, 620–625.
Suits, M. D., Zhu, Y., Taylor, E. J., Walton, J., Zechel, D. L., Gilbert, H. J. & Davies, G. J. (2010). *PLoS One*, **5**, e9006.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wiesmann, C., Beste, G., Hengstenberg, W. & Schulz, G. E. (1995). *Structure*, **3**, 961–968.
Wiesmann, C., Hengstenberg, W. & Schulz, G. E. (1997). *J. Mol. Biol.* **269**, 851–860.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
Wright, L. *et al.* (2004). *Chem. Biol.* **11**, 775–785.
Zechel, D. L., Boraston, A. B., Gloster, T., Boraston, C. M., Macdonald, J. M., Tilbrook, D. M. G., Stick, R. V. & Davies, G. J. (2003). *J. Am. Chem. Soc.* **125**, 14313–14323