Mycobacterium tuberculosis Strains Possess Functional Cellulases*

Received for publication, March 29, 2005, and in revised form, April 8, 2005
Published, JBC Papers in Press, April 8, 2005, DOI 10.1074/jbc.C500142200

Annabelle Varrot‡‡‡, Sabine Leydier‡‡‡, Gavin Poll¶¶, James M. Macdonald**, Robert V. Stick**, Bernard Henriassat‡‡, Harry J. Gilbert¶ and Gideon J. Davies$$

From the ¶¶York Structural Biology Laboratory, Department of Chemistry, University of York, Y010 5YW, United Kingdom, the §School for Cell and Molecular Biosciences, The Medical School, University of Newcastle upon Tyne, NE2 4HH, United Kingdom, ¶¶Chemistry, School of Biomedical and Chemical Sciences M313, University of Western Australia, Crawley, Western Australia 6009, Australia, and the §§Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS and Universités Aix-Marseille I & II, 31 Chemin Joseph Aiguier, 13402 Marseille cedex 20, France

The genomes of various Mycobacterium tuberculosis strains encode proteins that do not appear to play a role in the growth or survival of the bacterium in its mammalian host, including some implicated in plant cell wall breakdown. Here we show that M. tuberculosis H37Rv does indeed possess a functional cellulase. The x-ray crystal structure of this enzyme, in ligand complex forms, from 1.9 to 1.1Å resolution, reveals a highly conserved substrate-binding cleft, which affords similar, and unusual, distortion of the substrate at the catalytic center. The endoglucanase activity, together with the existence of a putative membrane-associated crystalline polysaccharide-binding protein, may reflect the ancestral soil origin of the Mycobacterium or hint at a previously unconsidered environmental niche.

Genome-wide inventory of the carbohydrate-active enzyme repertoire of mycobacteria reveals a significant number of open reading frames (ORFs) that encode proteins that do not play a role in the growth or survival of the bacterium in its mammalian host, including some apparently implicated in cellulose breakdown. For example, the Mycobacterium tuberculosis H37Rv genome, as well as those from M. tuberculosis CDC1551 and Mycobacterium bovis, appear to possess plant cell wall hydrolases from the sequence-based glycoside hydrolase (GH) families (2) GH5, GH6, GH12, and GH16. Indeed, many of the unfinished mycobacterial genomes (M. tuberculosis 210, Mycobacterium marinum M, Mycobacterium smegmatis str. MC2 155, and Mycobacterium avium 104) all possess similar ORFs, the functions of which are unknown.

M. tuberculosis H37Rv ORF Rv0062 encodes a protein, hereafter called Cel6, that contains an N-terminal sequence of 88 residues featuring an extended basic region and features consistent with a signal peptide, and a catalytic domain comprising residues 89–380, assigned to glycoside hydrolase family GH6. Other characterized GH6 members display two activities, termed cellobiohydrolases and endoglucanases, the former liberating the disaccharide cellobiose from crystalline cellulose, while the latter make random internal cuts on single β,1-4 glucan chains. Structurally, these two classes mainly differ in the size of the two extended loops forming the active site (3–5). The sequence alignment of M. tuberculosis H37Rv Cel6 with other GH6 members suggests that it could be an endoglucanase.

We have cloned and expressed the protein derived from M. tuberculosis H37Rv ORF Rv0062. Here we show that it is a functional cellulase that liberates a repertoire of cello-oligosaccharides from acid-swollen cellulose and barley β-glucan with an endo action. The x-ray crystal structure of this protein, solved in ligand-complexed forms (Fig. 1), reveals that Rv0062 encodes a functional, indeed typical, GH6 endoglucanase with a catalytic “−1” subsite (6) that promotes identical distortion of the substrate glucoside to that observed previously for homologs from plant cell wall-degrading bacteria and fungi (7). The observation that Cel6 displays the structure and biochemical properties of a “classical” GH6 endo-acting cellulase poses intriguing questions for its role in mammalian pathogenic mycobacteria.

MATERIALS AND METHODS

Bacterial Strains, DNA Manipulation, and Cloning—The Escherichia coli strains used in this study were XL2Blue and Origami (DE3) (Novagen), and the plasmids employed were pET22b and pET16b (Novagen) (8). E. coli was cultured in Luria broth supplemented with 15 mg l−1 kanamycin and tetracycline and 100 mg l−1 ampicillin at 37 °C to an A600 of ~0.6 before induction of Cel6 expression by the addition of 0.2 mM isopropyl 1-thio-β-D-galactopyranoside and incubation for a further 7 h. Cell-free extracts were prepared and His-Cel6 was purified by immobilized metal ion affinity chromatography using a Ni2+ -chelating column (Amersham Biosciences) and a 5–500 mM imidazole gradient in 20 mM Tris-HCl, pH 7.9, 0.5 M NaCl, to elute the recombinant protein. Fractions containing Cel6 (as judged by SDS-PAGE) were concentrated using a Vivaspin concentrator before further purification by gel filtration using a S75 16/26 column (Amersham Biosciences) equilibrated in 20 mM sodium HEPES buffer, pH 7.0.
pH 7.5, containing 100 mM NaCl, using an fast protein liquid chromatography system. The eluted protein was concentrated using a 30,000 cut-off filter, at 4 °C, and washed into water (MilliQ) to a final concentration of 10 mg ml⁻¹.

Cel6-native was purified by anion exchange chromatography using a DEAE column (Amersham Biosciences) using a 0–500 mM NaCl gradient. Frations containing Cel6, as judged by SDS-PAGE, were pooled, and ammonium sulfate was added to a final concentration of 1.4 M. After removal of insoluble material the protein solution was subjected to hydrophobic chromatography using a Phenyl-Sepharose column (Amersham Biosciences) and a 1.5 to 0 M ammonium sulfate gradient, in 20 mM Tris-HCl buffer, pH 8.0, to elute the protein. Cel6-native was further purified by gel filtration as described for His-Cel6.

Cel6 Assays—Purified Cel6 was assayed for activity against polysaccharides using the reducing sugar assay of Miller (9). Enzyme reactions were carried out in 50 mM sodium phosphate/12 mM citrate (PC) buffer, pH 6.5, containing 1 mg ml⁻¹ soluble polysaccharide or 4 mg ml⁻¹ insoluble polysaccharide and 1 mg ml⁻¹ bovine serum albumin. The reactions, which were incubated at 37 °C, were initiated by the addition of Cel6 to a final concentration of 500 nM. At regular intervals up to 20 min, 500-μl aliquots were removed and assayed for reducing sugar (9), and the product profile was determined by high pressure anion exchange chromatography (HPAEC) (10). To evaluate the reaction products, aliquots were incubated for 30 min prior to crystallization. The complex of Cel6-native with thiocellulose (SDP5) was crystallized using 12–14% (w/v) PEG 4000 and 200 mM lithium acetate, pH 6.5, containing 1 mg ml⁻¹ bovine serum albumin. The reactions, which were incubated at 37 °C, were initiated by the addition of Cel6 to a final concentration of 500 nM. At regular intervals up to 20 min, 500-μl aliquots were removed and assayed for reducing sugar (9), and the product profile was determined by high pressure anion exchange chromatography (HPAEC).

Crystallization and Structure Solution—All crystals were obtained by hanging-drop vapor diffusion with 1-μl drops containing 50:50 (v/v) mix of the protein and the reservoir solution. The protein was incubated with 1 mM ligand for at least 30 min prior to crystallization. The complex of Cel6A-native with thiocellulose (SDP5) was crystallized using 12–14% (w/v) PEG 4000 and 200 mM lithium sulfate, as precipitant, in 100 mM MES buffer, pH 6.5. Cells with cellulobio-derived isofagomine (ISOF) were obtained using 12–14% (w/v) PEG 4000 and 200 mM lithium sulfate, as precipitant, in 100 mM sodium acetate buffer, pH 4.6. Cryoprotection involved supplementing the growth conditions with 20% (v/v) glycerol or 30% PEG 400, respectively. Data for His-Cel6-SDP5 were collected at European Synchrotron Radiation Facility ID14–2 and beamline Cel6-ISOF in the home laboratory using a CuKa rotating-anode source operating at 50 kV and 100 mA, with Osmic mirrors and a MAR research detector. Data were processed with MOSFLM from the CCP4 suite (11) or the HKL suite of programs (12). All further computing was performed using the CCP4 suite, unless otherwise stated.

The structure of the His-Cel6-SDP5 was solved by molecular replacement using the coordinates from the endoglucanase from Thermobifida fusca (Protein Data Bank code 1TML) as a search model. The program AMoRe (13) was used in conjunction with data in the resolution range 20–4 Å. This structure was subsequently used to solve the other Cel6 complexes. 5% of the observations was set aside for cross-validation analysis (14) and was used to monitor various refinement strategies with REFMAC (15). Manual corrections of the model were performed using the X-FIT routines of the program QUANTA (Accelrys, San Diego, CA). Water molecules were added in an automated manner using ARP (16) and verified manually. The crystal structures of Cel6 with cellulobiose (Protein Data Bank code 1UP0) and cellulose-β-cellobiose (Protein Data Bank code 1UF3) were also solved at resolutions of 1.75 and 1.6 Å and deposited but are not discussed further.

RESULTS AND DISCUSSION

The Three-dimensional Structure of Cel6—Cel6 was crystallized in various ligand-bound forms, with non-hydrolyzable thio-oligosaccharides and with a disaccharide inhibitor, cellulobio-derived isofagomine (Fig. 1) and data collected between 1.9 and 1.1 Å resolution (Table 1). The catalytic domain presents a distorted (a/b)₈ barrel typical for family GH6 enzymes (3–5) (Fig. 2a) composed of eight β-strands, with seven involved in the β-barrel formation and eight α-helices. As predicted from sequence, the structure possesses the shortened C-terminal loop typical of endoglucanases, consistent with its biochemical properties (see below). In marked contrast to previous GH6 endoglucanase structures, however, the active-center loops of Cel6 enclose the ligand-binding cleft (Fig. 2a) rather than remaining “open.” Complexes with thio-oligosaccharides reveal binding in the aglycone (leaving group) subsites, termed +1 and +2, with sugars in the 4C₁ (chair) conformation (data not shown). A “kink” at +2 may allow for the formation of a third “4-3” subsite (and there is ample protein surface at the non-reducing end of the −2 site to accommodate binding of longer substrates). A strong indication that this mycobacterial enzyme is a functional β-glucanase is that the complex structures confirm that all the direct interactions of the “glycone” −2 and −1 subsites and aglycone +1 sites are invariant. The +2 subsite shows more divergence but still maintains its hydrophobic sugar-binding platform, Trp250.

Substrate Binding and Distortion at the Catalytic Center—Further evidence for the catalytic integrity of Cel6 comes from its

![Fig. 1. Oligosaccharides used in the structure determination of the M. tuberculosis Cel6. Compound 1 thio-cellobiose (20) (SDP5) and compound 2 a cellulobio-derived isofagomine (21) are shown.](http://www.jbc.org/content/11001-11002/fig1)

![Fig. 2. a, three-dimensional structure of the cellulobio-derived isofagomine complexes of the M. tuberculosis Cel6 (green, this study) overlaid with the plant cell wall-degrading endoglucanase Cel6B from H. insolens (red) (22). b, maximum likelihood weighted 2mFo−DFc electron density (contoured at 1 σ, 0.44 eÅ⁻³) for the cellulobio-derived isofagomine complex of Cel6, in which two molecules of the disaccharide are bound in the −2/−1 and +1/+2 subsites, respectively. Residues mentioned in the text are shown. These figures, in divergent (wall-eyed) stereo, were drawn with BOBSCRIPT (23).](http://www.jbc.org/content/11001-11002/fig2)
complex with the cellobio-derived isofagomine. This is a disaccharide inhibitor in which a glucosyl moiety is β-1,4 linked to the inhibitor isofagomine (possessing a nitrogen group in place of the anomic carbon (Fig. 1)). It is presumed to be a tight-binding inhibitor by virtue of the transition state-mimicking positive charge on the ring nitrogen of the compound as its conjugate acid (7). In the Cel6 structure, two molecules of the inhibitor occupy the substrate-binding site from subsite −2 to −1 and +1 to +2 (Fig. 2b). In subsites −2, +1, and +2 the glucoside and isofagomine rings lie in relaxed C1′ (chair) conformation, while in the “catalytic” −1 subsite the aza-sugar moiety is distorted to a conformation intermediate between S2′ (skew-boat) and S2′ (boat) conformation, as observed previously for the cellohydrase Cel6A from Humicola insolens (7). The S2′ and S2′ are implicated as the transition state conformation during catalysis, as it maintains C5, O5, C1, and C2 co-planar, an essential prerequisite of an oxocarbonium-ion-like species. Notably, the active center of Cel6 possesses the constellation of aspartates required for transition state stabilization, distortion, and catalysis (3, 7): Asp168, Asp353, and the catalytic acid, Asp206, and all the usual substrate-binding residues that flank the active center (Fig. 3). The sole difference in the catalytic subsite of Cel6, compared with the previously published GH6 enzymes, is that a Ser residue implicated in stabilizing a solvent water in an unusual "Grothhus" inverting mechanism is replaced by Ala. The putative attacking water, however, still resides "below" the anomic carbon (Figs. 2b and 3).

**Catalytic Activity of Cel6**—Demonstration of the catalytic function of the recombinant enzyme is shown by its catalytic activity on both barley β-glucan and acid swollen cellulose. The enzyme hydrolyzes both substrates to release an array of different gluco-oligosaccharides, a product profile typical of endo-acting β-glucanases (Fig. 4). Reducing sugar analysis of the reaction products showed that Cel6 displayed turnover rates (mole of product per mole of enzyme per second) of 25 s⁻¹ and 0.9 s⁻¹ on β-glucan and acid-swollen cellulose, respectively. These are of the same order as observed for endo-acting GH6 cellulases from known saprophytic plant cell wall-degrading microorganisms that are typically in the range 1–10 s⁻¹ on acid-swollen cellulose, for example. The enzyme displays no activity against a range of other polysaccharides including amylose, xylan, mannan, arabanin, pectin, laminarin, and a range of aryl glucosides including 4-nitrophenyl β-D-celllobioside. Cel6 is, however, also able to hydrolyze cellulitooligosaccharides; cellobentaose is converted into cellobiose and cellotriose; cellobiohexaose into cellobiose, cellotriose, and cellotetraose in a molar ratio of 1:4:1 (data not shown), while the enzyme displays no measurable activity against cellotetraose or cellotriose. Thus, at least five subsites in Cel6, as also predicted from the crystal structures of the enzyme-ligand complexes, make a highly significant contribution to catalysis.

What are the implications for *M. tuberculosis*, a non-plant cell wall-degrading organism, possessing functional cellulases? That they are most likely important is underlined by the fact that all finished mycobacterial genomes, excluding *Mycobacterium leprae* (discussed below) together with *M. tuberculosis* 210; *M. marinus* 104, and *M. smegmatis* str. MC2 155 possess a GH6 cellulase, at least one GH3 β-glucosidase candidate (the enzyme that is required for the subsequent degradation of cellulase-derived products), and potential β-glucosidases from GH5 and GH16. *M. smegmatis* str. MC2 155 also possesses a candidate β-glucosidase from family GH1. It is interesting to note that the GH5 and GH16 enzymes form distinct “subfamilies” when compared with the other respective characterized family members and one cannot speculate on their function beyond saying that they will be retaining β-glucosidases with potential for transglycosylation.

Despite possessing the enzymes responsible for cellulose degradation, it is, however, extremely unlikely that *M. tuberculosis*, or related mycobacterial strains, can utilize the plant cell wall as a significant nutrient as they lack the full repertoire of hydrolytic enzymes required to attack this composite structure. For example, no xylologlucanase, α-glucuronidase, or xylanase candidates are found in any of the mycobacterial genomes. That Cel6 possesses a signal peptide certainly points to an extracellular role. Equally important is the sole obligate parasite among mycobacteria, *M. leprae*, possesses none of the potential cellulase candidates on its genome.
It is possible that these enzymes date from a prehistoric time when primitive mycobacteria were soil-based organisms, as M. bovis is still. Indeed Cel6 displays significant sequence identity (~40%) with numerous Streptomyces GH6s indicating a strong evolutionary relationship among these enzymes (Fig. 5). This is consistent with the view that Streptomyces and Mycobacterium are related Actinomycetes. It is estimated that these bacteria diverged from a common ancestral organism around 80 million years ago. It is possible, therefore, that Cel6 is derived from the ancestral organism. The striking similarity between specific islands within the genomes of Streptomyces coelicolor A3 (2) and M. tuberculosis, however, suggests that gene transfer has occurred between the organisms when the bacteria occupied soil niches, and these events could also have resulted in the acquisition of the cellulase gene by the Mycobacterium. Irrespective of the mechanism of Cel6 acquisition by M. tuberculosis, it remains difficult to provide definitive insight into its role in the human pathogen. It is possible that the gene is an evolutionary relic and there has not been sufficient evolutionary time for its removal. Yet, it is surely significant that while M. tuberculosis only contains β-glucanases and a cellulose-binding module, Streptomyces coelicolor A3 contains the complete and extensive repertoire of hydrolytic enzyme genes that encode proteins that attack all the major plant structural polysaccharides. Thus, the retention of a small number of cellulase genes within the genome of the Mycobacterium suggests that these genes do confer a selective advantage to the organism. Is it possible that there is a previously undescribed secondary host such as a rumen protozoan. Such a species would possess a rudimentary plant cell wall-degrading system, in which cellulose, or its degradation products, could be presented to the Mycobacterium. An alternative evolutionary rationale for the β-glucanase activity displayed by M. tuberculosis is that the bacterium encounters biofilms and that some of these exo-poly saccharides on which these tight microbial ecosystems form are β-glucans. While the chemical composition of most extracellular polysaccharides that are integral to biofilm formation is unknown, some, such as that made by Pseudomonas aeruginosa, have recently been shown to consist of β-glucans (17). Indeed, a recent report has identified the region in the important human pathogen P. aeruginosa (18) that enables the bacterium to make this switch between a virulent, disease-causing state and a biofilm state in the mammalian host. Of interest in this context is the occurrence of a gene encoding a membrane-associated family 2 carbohydrate-binding module, CBM2, (CBM2s are known to bind to β-glucans such as cellulose and chitin, reviewed in Ref. 19), yet another incongruous open reading frame in all completed, and four unfinished, mycobacterial genomes.

Acknowledgments—We thank Hugues Drigué (CNRS, Grenoble, France) for the kind provision of thio-oligosaccharide ligands and the staff of the European Synchrotron Radiation Facility (Grenoble, France) for provision of data collection facilities.

REFERENCES

1. Cole, S. T., Cole, S. T., Barksch, R., Parkhill, J., Garnier, T., Church, C., Harris, D., Gordon, S. V., Kigame, K., Gas, S., Barry, C. E. III, Tekaija, P., Badock, K., Basham, D., Brown, D., Chillingworth, T., Conner, R., Davies, R., Devlin, K., Fellet, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krugh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osbourne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrett, B. G. (1998) Nature 393, 537–544

2. Coutinho, P. M., and Henriassett, B. (1995) in Recent Advances in Carbohydrate Engineering (Gilbert, H. J., Davies, G. J., Svensson, B., and Henriassett, B., eds) pp. 3–12, Royal Society of Chemistry, Cambridge, UK

3. Krouvines, J., Bergfors, T., Teeri, T., Knowles, J. K., and Jones, T. A. (1999) Science 249, 380–386

4. Spezio, M., Wilson, D. B., and Karplus, P. A. (1993) Biochemistry 32, 9906–9916

5. Varrot, A., Schulien, M., and Davies, G. J. (1999) Biochemistry 38, 8884–8891

6. Davies, G. J., Wilson, K. S., and Henriassett, B. (1997) Biochem. J. 321, 557–559

7. Varrot, A., Macdonald, J., Stick, R. V., Pell, G., Gilbert, H. J., and Davies, G. J. (2003) Chem. Commun. 946–947

8. Studier, F. W., and Moffatt, B. A. (1988) J. Mol. Biol. 189, 113–130

9. Miller, G. L. (1959) Anal. Chem. 31, 426–428

10. Hall, J., Black, G. W., Ferreira, L. M., Millward-Sadler, S. J., Alt, B. R., Hazelwood, G. P., and Gilbert, H. J. (1995) Biochem. J. 309, 749–756

11. Collaborative Computational Project Number 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763

12. Otwinowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326

13. Varona, J., and Sabudjian, P. (1997) Methods Enzymol. 276, 581–594

14. Brünger, A. T. (1992) Nature 355, 472–475

15. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255

16. Lammaz, V. S., and Wilson, K. S. (1993) Acta Crystallogr. Sect. D Biol. Crystallogr. 49, 129–147

17. Mah, T. F., Pitts, B., Pellock, B., Walker, G. C., Stewart, P. S., and O’Teagle, G. A. (2003) Nature 426, 306–310

18. Goodman, A. L., Kulasekara, B., Rietse, A., Boyd, D., Smith, R. S., and Lory, S. (2004) Dev. Cell 7, 745–754

19. Poratson, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Biochem. J. 382, 769–781

20. Schou, C., Rasmussen, G., Schulein, M., Henriassett, B., and Drigué, H. (1993) J. Carbohydr. Chem. 12, 743–752

21. Maechold, J., Stiek, H. V., Tilbrook, D. M. G., and Withers, S. (2002) Aust. J. Chem. 55, 747–752

22. Davies, G. J., Brozowski, A. M., Dauter, M., Varrot, A., and Schulien, M. (2000) Biochem. J. 348, 291–297

23. Eronen, R. M. (1997) J. Mol. Graph. Model. 15, 132–134

24. Edgar, R. C. (2004) BMC Bioinformatics 5, 113

25. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673–4680

26. Page, R. D. M. (1996) Comput. Appl. Biosci. 12, 357–358
