Mycobacterium tuberculosis H37Rv3377c encodes the diterpene cyclase for producing the halimane skeleton†

Chiaki Nakano,[a] Tomoo Okamura,[a] Tsutomu Sato,[a] Tohru Dairi[b] and Tsutomu Hoshino*[a]

Received (in Cambridge, UK) 4th October 2004, Accepted 15th November 2004
First published as an Advance Article on the web 4th January 2005
DOI: 10.1039/b415346d

The cloning and functional expression of *Mycobacterium tuberculosis* Rv3377c in *Escherichia coli* revealed that this gene encodes the diterpene cyclase for producing (+)-5(6),13-halimadiene-15-ol, which accepts geranylgeranyldiphosphate as the intrinsic substrate.

Despite over a century of research, *Mycobacterium tuberculosis* (MT) remains a leading cause of infectious death worldwide. Since the genome project finished in 1998,[1] numerous studies have appeared to elucidate the biochemical reaction of each open reading frame. For example, sterol 14α-demethylase P450 is demonstrated to be involved in MT H37Rv,[2] suggesting the possibility of cholesterol biosynthesis by MT. Supportive evidence for the ability of mycobacteria to synthesize cholesterol has been given; a trace amount of cholesterol was recently reported to be found in *M. smegmatis*.[3] MT H37Rv3377c protein, consisting of 501 amino acid residues,[1,4] has two characteristic alignments, called QW[5] and DXDD motifs,[5] which are also conserved in oxidosqualene and squalene cyclases and/or other terpenoid cyclases.[5] To identify the biochemical function, the Rv3377c was cloned, functionally expressed in *E. coli*, and the expressed protein was subjected to the enzymic reactions using various substrates of terpenoid cyclases. We present here the definitive evidence that Rv3377c gene product encodes the diterpene cyclase to produce the halimane skeleton. It should be noted that this gene is involved only in the pathogenic species of mycobacteria.

The Rv3377c gene was amplified by PCR and cloned into the *Bam*H I/*Hin*d III site of pET-22b(+). The recombinant enzyme was expressed in *E. coli* BL21(DE3) under the following conditions: isopropylthiogalactoside, 0.4 mM; induction for 10 h at 20 °C, but almost all of the expressed protein formed an inclusion body. A coproduction of the Rv3377c and GroE chaperon[6] in *E. coli* was only slightly effective for solubilizing the recombinant protein (at most *ca.* 5%), but highly effective for the enzyme activity due to the correct folding of the Rv3377c enzyme; no enzymic activity was found unless the chaperon was coexpressed. The following materials were tested as the substrates: geranyl- (GPP), farnesyldiphosphate (FPP), geranylgeranyldiphosphate (GGPP, **1**), squalene and (3*R*,*S*)-2,3-oxidosqualene. These substrates were incubated at 25 °C and pH 7.5[7] for 20 h with the cell-free homogenates containing Triton X-100 (1%), dithiothreitol (1 mM) and $MgCl_2$ (10 mM), which were prepared from the recombinant *E. coli* having the chaperon-coexpression system. GC analyses of the hexane extract from the reaction mixtures showed that no conversion was found for GPP, FPP, squalene and (3*R*,*S*)-oxidosqualene, but **1** was converted in a high yield (*ca.* 50%), indicating that Rv3377c encodes a diterpene synthase.

To isolate a sufficient amount of the enzymic product **2** for the structure determination, 8 mg of **1** were incubated with the cell-free homogenates from 5-L culture. The GC trace showed the presence of both geranylgeraniol (GGOH) and **2** in addition to Triton X-100 (Fig. 1A). The hexane extract from the incubation mixture was subjected to SiO_2 column chromatography eluting with hexane–EtOAc (100 : 2), yielding a pure product **2** (3 mg). The structure was determined to be 5(6),13-halimadiene-15-ol by the detailed 2D NMR analyses (Scheme 1). The relative stereochemistry of **2** was established by the NOESY spectrum. We propose to name **2** tuberculosinol. Product **2** had a primary alcoholic group (δ_H 4.10, 2H, d, *J* 6.5, H-15; δ_C 59.4, t, C-15) and no signal in the ^{31}P NMR spectrum, suggesting that the diphosphate group of the enzymic product was eliminated during the enzymic reaction. Addition of phosphatase inhibitor[8] led to no recovery of **2** from the hexane extract, indicating that the polar diphosphoryl group of the enzymic product had been eliminated by action of the endogenous phosphatase of *E. coli*. In addition, from the incubation mixture of **1** with the cell-free extract, which was prepared from *E. coli* lacking the Rv3377c gene, GGOH was also found in the hexane extract (Fig. 1B), but addition of the phosphatase inhibitor[8] led to no recovery of GGOH. These findings indicated that Rv3377c enzyme itself had no phosphatase activity and catalyzed only the reaction of **1** → **4**. The cyclization of **1** proceeded in a pre chair–chair conformation to give bicyclic **3** having a C8-cation. A series of 1,2-shift reactions of hydride and

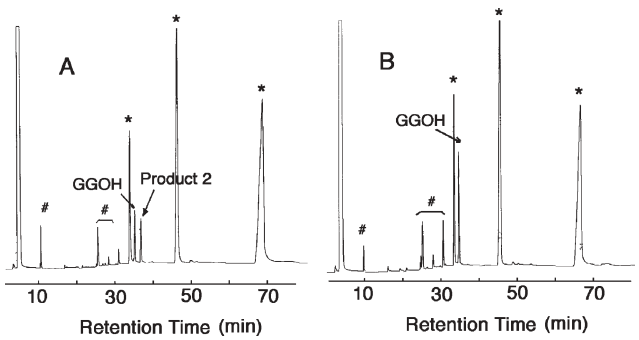
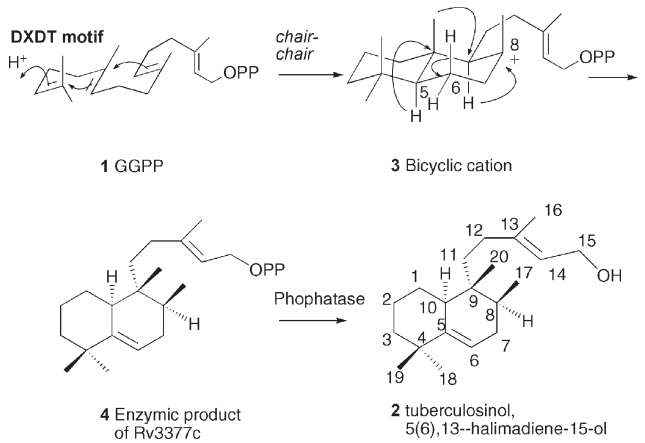


Fig. 1 GC traces of the hexane extracts obtained by incubating **1** with the cell-free extracts from *E. coli* having Rv3377c (**A**) and lacking this gene (**B**). The GroE-coexpression system was used. The symbols * and # show the Triton X-100 and impurities, respectively.

† Electronic supplementary information (ESI) available: cloning and expression protocols of Rv3377c, amino acid alignment of some diterpene cyclases, primers for site-directed mutants, and the enzyme purification method. See http://www.rsc.org/suppdata/cc/b4/b415346d/

*hoshitsu@agr.niigata-u.ac.jp



Scheme 1 Proposed cyclization mechanism of **1** into **4** by Rv3377c cyclase. The diphosphoryl group was hydrolyzed by endogenous phosphatase of *E. coli*. The depicted structure of **2** shows the relative stereochemistry.

methyl group in antiparallel fashion, followed by deprotonation of H-6, introduced the double bond at C5–C6 to give **4** (Scheme 1).

Two characteristic motifs of DXDD and DDXXD are usually found in diterpene cyclases. It is generally agreed that the DXDD motif affords the acidic proton on the terminal double bond of linear substrates to initiate cyclization reactions, leading to the production of cyclic diterpenes.[5,9,10] Comparison of amino acid alignment[10] between Rv3377c and other diterpene cyclases, *e.g.* abietadiene and copalyldiphosphate synthases, strongly suggests that the $DXDT_{293–296}$ sequence of Rv3377c corresponds to the DXDD motif.[11] As anticipated, point mutants of D293→N and D295→N gave no cyclization product, because of a decrease in the acidity of the motif, but that of T296→D also unexpectedly resulted in the loss of the cyclase activity, despite the acidity of the DXDT motif having been increased. On the other hand, the DDXXD motif[11] is known to have a crucial role in the cyclization reactions initiated by the release of the diphosphate group.[11] This motif is missing in Rv3377c,[10] thus, **1** could be cyclized according to Scheme 1 (protonation-initiated cyclization). A divalent cation of Mg^{2+} was demonstrated to be essential to the cyclization reaction; the homogeneous enzyme, which was purified in the presence of EDTA, had no cyclization activity, but supplementation of $MgCl_2$ (1 mM) afforded **2** in a high yield (94.8%).[12] The diphosphate moiety of **1** may bind to the cyclase through the chelation of Mg^{2+} cation.[9,11] The essentiality of the diphosphate group for the cyclization reaction was further supported by the fact that GGOH itself was not cyclized. Further studies on the cyclase are underway.

In conclusion, Rv3377c was demonstrated to encode a diterpene cyclase to give the halimane skeleton. Production of diterpene by prokaryotes is very rare. To date, only one example of *Kitasatospora griseola* (formerly named *Streptomyces griseolosporeus*) has been reported, which produces an antibiotic having a clerodane diterpene skeleton, called terpentecin.[13] Production of the halimane skeleton by prokaryotes has not hitherto been reported. A halimane skeleton is usually found in eukaryotes such as plants and liverworts,[14] suggesting that this gene might have been acquired by horizontal gene transfer from eukaryotes, as suggested for sterol 14α-demethylase and the related sterol-metabolizing enzymes found in MT.[15] Insight into the biological function(s) of this halimane product is of particular interest, especially from the viewpoint of the pathogenicity of MT. The same gene as Rv3377c is also found in other pathogenic species (*M. tuberculosis* CDC1551[16] and *M. bovis subsp. bovis* AF2122/97[16,17]), but is not found in non-pathogens (*M. smegmatis* MC2155,[18] *M. avium*104,[18] and *M. avium subsp. oaratuberculosis* str.k10[16]). This fact strongly suggests that Rv3377c may closely relate to the pathogenicity.[19]‡

We thank Professor Ishii, University of Tsukuba, Japan, who kindly provided the plasmid pT-groE (GroESL expression vector).[6] This work was made possible by the financial support (Nos. 15380081, 15780083 and 16208012) provided by the Ministry of Education, Science and Culture of Japan.

Chiaki Nakano,[a] Tomoo Okamura,[a] Tsutomu Sato,[a] Tohru Dairi[b] and Tsutomu Hoshino*[a]
[a]Department of Applied Biological Chemistry, Faculty of Agriculture, and Graduate School of Science and Technology, Niigata University, Ikarashi, Niigata, 950-2181, Japan.
E-mail: hoshitsu@agr.niigata-u.ac.jp; Fax: +81-25-262-6854
[b]Biotechnology Research Center, Toyama Prefectural University, Toyama, 939-0398, Japan

Notes and references

‡ Product **2**. NMR data in C_6D_6: δ_H 0.81 (3H, s, H-20), 0.92 (3H, d, *J* 6.8, H-17), 1.21 (m, H-1), 1.22 (3H, s, H-19), 1.27 (3H, s, H-18), 1.38 (m, H-3), 1.50 (m, H-11), 1.52 (m, H-3), 1.63 (3H, s, H-16), 1.64 (m, H-8), 1.66 (m, H-11), 1.67 (2H, m, H-2), 1.86 (m, H-1), 1.94 (2H, m, H-7), 1.99 (ddd, *J* 4, 12.8, 13.2, H-12), 2.06 (ddd, *J* 4, 12.8, 13.2, H-12), 2.37 (bd, *J* 11.8, H-10), 4.10 (2H, d, *J* 6.5, H-15), 5.54 (bt, *J* 6.7, H-14), 5.67 (m, H-6); δ_C 15.28 (C-17), 16.39 (C-16), 16.43 (C-20), 22.60 (C-2), 27.22 (C-1), 29.15 (C-19), 30.04 (C-18), 31.97 (C-7), 33.05 (C-12), 33.70 (C-8), 35.35 (C-11), 36.28 (C-4), 37.25 (C-9), 40.31 (C-10), 41.25 (C-3), 59.41 (C-15), 116.73 (C-6), 124.5 (C-14), 139.2 (C-13), 146.2 (C-5). The assignments of H-18 and H-19, those of C-18 and C-19, and those of C-16 and C-20 may be exchangeable. EIMS (%) *m/z* 69 (31), 80 (100), 93 (37), 119 (80), 136 (55), 189 (81), 190 (73), 191 (68), 290 (11). HRMS (EI) calcd. for $C_{20}H_{34}O$, 290.2610; found 290.2639. $[\alpha]_D^{25}$ + 22.2 (*c* 0.3, EtOH).

1 S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead and B. G. Barrell, *Nature*, 1998, **393**, 537–544.
2 (*a*) A. Bellamine, A. T. Mangla, W. D. Nes and M. R. Waterman, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 8937–8942; (*b*) Y. Yoshida, M. Noshiro, Y. Aoyama, T. Kawamoto, T. Horiuchi and O. Gotoh, *J. Biochem.*, 1997, **122**, 1122–1128.
3 D. C. Lamba, D. E. Kellya, N. J. Manning and S. L. Kellya, *FEBS Lett.*, 1998, **437**, 142–144.
4 *Mycobacterium tuberculosis* Structural Genomics Corsortium. http://www.doe-mbi.ucla.edu/TB.
5 (*a*) T. Hoshino and T. Sato, *Chem. Commun.*, 2002, 291–301; (*b*) T. Sato and T. Hoshino, *Biosci. Biotechnol. Biochem.*, 1999, **63**, 2189–2198; (*c*) K. U. Wendt, G. E. Schulz, E. J. Corey and D. R. Liu, *Angew. Chem., Int. Ed.*, 2000, **39**, 2812–2833.
6 T. Yasukawa, C. Kanei-Ishii, T. Maekawa, J. Fujimoto, T. Yamamoto and S. Ishii, *J. Biol. Chem.*, 1995, **270**, 25328–25331.
7 The optimum catalytic pH and temperature for **1** → **2** were 6.5–7.5 and 15–25 °C, respectively; see ESI.
8 Phosphatase inhibitors of cocktail 1 and 2 from SIGMA Co. were tested. Addition of the cocktail 1 led to the recoveries of GGOH and **2** from the hexane extract of the incubation mixture, but that of

cocktail 2 resulted in no recovery of these alcohols, indicating that the cocktail 2 effectively inhibited the phosphatase activity of GGPP to GGOH and that of **4** to **2**.
9 (*a*) R. J. Peters and R. B. Croteau, *Biochemistry*, 2002, **41**, 1836–1842; (*b*) R. J. Peters, M. M. Ravn, R. M. Coates and R. B. Croteau, *J. Am. Chem. Soc.*, 2001, **123**, 8974–8978; (*c*) H. Kawaide, T. Sassa and Y. Kamiya, *J. Biol. Chem.*, 2000, **275**, 2276–2280.
10 Amino acid alignment of a few diterpene cyclases is given in the ESI.
11 J. C. Sacchettini and C. D. Poulter, *Science*, 1997, **277**, 1788–1789.
12 No detectable amount of **2** was found in the hexane extract from the incubation mixture when **1** was incubated with the homogeneously purified Rv3377c enzyme, further supporting the proposal that this enzyme catalyses only the reaction of **1** → **4**. To eliminate the diphosphate group of **4** to afford **2**, acid phosphatase was added to the preincubation mixture of **1** with the purified Rv3377c enzyme, followed by further incubation; see, H. Fujii, T. Koyama and K. Ogura, *Biochim. Biophys. Acta*, 1982, **712**, 716–718.
13 Y. Hamano, T. Kuzuyama, N. Itoh, K. Furihata, H. Seto and T. Dairi, *J. Biol. Chem.*, 2002, **277**, 37098–37104.
14 (*a*) M. Abdel-Kader, J. M. Berger, C. Slebodnick, J. Hoch, S. Malone, J. H. Wisse, M. C. W. Werhoven, S. Mamber and D. G. I. Kingston, *J. Nat. Prod.*, 2002, **65**, 11–15; (*b*) N. Hara, H. Asaki, Y. Fujimoto, Y. K. Gupta, A. K. Singh and M. Sahai, *Phytochemistry*, 1995, **38**, 189–194.
15 J. Gamieldien, A. Ptitsyn and W. Hide, *Trends Genetics*, 2002, **18**, 5–8.
16 DNA Data Bank of Japan (DDBJ); http://www.ddbj.nig.ac.jp/.
17 T. Garnier, K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon and R. G. Hewinson, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 7877–7882.
18 The Institute for Genomic Research (TIGR); http://www.tigr.org/.
19 Very recently, Russell *et al.* suggested the horizontal gene transfer of Rv3377c and they constructed the transposon-inserted mutants that are defective in arrest of phagosome maturation, in which Rv3377c gene had been disrupted. See, K. Pethe, D. L. Swenson, S. Alonso, J. Anderson, C. Wang and D. G. Russell, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 13642–13647.

