ELSEVIER

Carbohydrate Research 276 (1995) 449–455

CARBOHYDRATE RESEARCH

Note

Structural definition of the glycopeptidolipids and the pyruvylated, glycosylated acyltrehalose from *Mycobacterium butyricum*

Kay-Hooi Khoo [a,1], Russell Suzuki [b], Howard R. Morris [a], Anne Dell [a], Patrick J. Brennan [b], Gurdyal S. Besra [b,*]

[a] *Department of Biochemistry, Imperial College of Science, Technology and Medicine, London SW7 2AY, UK*
[b] *Department of Microbiology, Colorado State University, Fort Collins, CO 80523, USA*

Received 20 January 1995; accepted in revised form 28 April 1995

Keywords: Glycopeptidolipids; Pyruvylated, glycosylated acyltrehalose; *Mycobacterium butyricum*; Structural analysis

Refined analytical techniques coupled with modern instrumentation have continued to furnish detailed structural knowledge on the species- and serotype-specific glycolipids of mycobacteria. These highly diverse glycosylated surface antigens have been broadly classified into three generic structures, namely the phenolic glycolipids (PGL), the glycopeptidolipids (GPL), and the trehalose-containing lipooligosaccharides (LOS) [1,2]. Using a combination of chemical, high-resolution NMR, and FABMS analyses, we and others have recently shown that the GPL family can be expanded beyond the "C-mycoside" GPL class to include those containing *O*-methylserine, i.e., the serine-containing GPLs of *Mycobacterium xenopi* [3–5]. These glycolipids often constitute the chemical basis for the identification and characterization of some atypical mycobacteria which cause opportunistic infections in AIDS [6]. Also, they are the basis of novel genetic approaches to effect the transformation of one mycobacterium with the glycolipids of another [7] for purposes of probing their roles in the pathogenesis of mycobacteria. In an effort to extend the study of these products, we were intrigued by the reported presence of an *N*-methyl-*O*-methylserine residue within the C-mycoside GPL from *M. butyricum* in the earlier literature [8], which prompted us to re-examine the surface glycolipids

[1] Present address: Department of Microbiology, Colorado State University, Fort Collins, CO 80523, USA.
[*] Corresponding author.

SSDI 0008-6215(95)00184-0

from this species using the FABMS methodology which we have developed for defining the structures of GPLs [3,9] and LOS [10].

The washed lipid fraction obtained from both *M. butyricum* ATCC 19979 and 559, when solubilized using hot acetone [3] and chromatographed on a Florisil column irrigated with $CHCl_3$ followed by gradual increments of CH_3OH in $CHCl_3$ [3], yielded two sets of partially purified glycolipids. The apolar set afforded a distinct yellow coloration when charred with 10% sulfuric acid in ethanol and, when treated with alkali, was found to be still lipid-soluble, features indicative of the alkali-stable GPLs [2]. The predominant component of the more polar glycolipid fraction, which gave a black color upon charring, co-chromatographed with the previously characterized glycosylated acyltrehalose from a phage resistant strain of *M. smegmatis* $mc^2 11$ [10].

The partially purified apolar GPLs were pooled and treated with mild alkali. The resultant deacylated GPLs were further purified by preparative thin-layer chromatography (TLC) using $CHCl_3$–CH_3OH–H_2O (90:10:1, v/v/v) to yield dGPL-I (3.4 mg, R_f 0.50), dGPL-II (4.1 mg, R_f 0.62), dGPL-III (2.0 mg, R_f 0.71), and dGPL-IV (5.0 mg, R_f 0.82). The nature of the individual glycosyl units in all four dGPLs was examined through hydrolysis, reduction with $NaBD_4$, per-*O*-acetylation, GLC, and GLC–MS of the alditol acetates as described [9]. GLC of the alditol acetates of dGPL-I and -II showed the presence of two glycosyl residues in equimolar ratio. Co-chromatography with 1,2,5-tri-*O*-acetyl-3,4-di-*O*-methylrhamnitol and 1,2,3,4,5-penta-*O*-acetyl-6-deoxytalitol together with GLC–MS analysis firmly established these residues as 3,4-di-*O*-methyl-Rha*p* (t_R 5.98 min; *m/z* 130, 131, 190) and 6-deoxy-Tal*p* (t_R 6.70 min; *m/z* 129, 171, 217, 231), respectively. In contrast, dGPL-III and -IV differed from dGPL-I and -II in containing a 2,3,4-tri-*O*-methyl-Rha*p* (t_R 5.45 min; *m/z* 115, 118, 131, 162, 175) replacing the 3,4-di-*O*-methyl-Rha*p* residue. All three glycosyl residues were determined to be in the L absolute configuration by comparative GLC–MS analyses of trimethylsilylated (*S*)-(+)-*sec*-butyl glycosides and (*R*)-(−)-*sec*-butyl-glycosides prepared from authentic standards [11]. Similar analyses on the *N*(*O*)-heptafluorobutyryl butyl esters of the amino acids [12] of dGPL-I–IV established their identities as D-Phe, D-allo-Thr, D-Ala, and alaninol.

The ^1H NMR spectra of dGPL-I and -II revealed the presence of two anomeric protons centered at 4.90 ppm (broad singlet, $^1J_{C1,H1}$ 168 Hz) and 4.70 ppm (doublet, $^3J_{H1,H2} < 1.5$ Hz, $^1J_{C1,H1}$ 168 Hz), which were assigned to 6-deoxy-α-L-Tal*p* and 3,4-di-*O*-methyl-α-L-Rha*p*, respectively. Resonances associated with the two methoxy groups of the 3,4-di-*O*-methyl-α-L-Rha*p* residue (3.29 ppm, singlet; 3.39 ppm, singlet) were also apparent. Similarly, the ^1H NMR spectra of dGPL-III and -IV revealed the characteristic anomeric resonances of 6-deoxy-α-L-Tal*p* at 4.90 ppm and 2,3,4-tri-*O*-methyl-α-L-Rha*p* at 4.78 ppm (doublet, $^3J_{H1,H2} < 1.5$ Hz, $^1J_{C1,H1}$ 168 Hz), and the methoxy resonances of the 2,3,4-tri-*O*-methyl-α-L-Rha*p* at 3.49, 3.43 and 3.41 ppm. In addition, dGPL-II and -IV possessed an extra methoxy singlet at 3.23 ppm which may be associated with the *N*-acyl chain. Thus, it appears that dGPL-I–IV are closely related, with the major distinctions being attributed to the total number and location of the methoxy substitutents, which would account for their different migration patterns on TLC.

The underivatized dGPLs each afforded two major ions in their FAB mass spectra

Table 1
The m/z values of the major ions afforded by positive-ion FAB mass spectra of perdeuteromethylated dGPLs

Signal assignment [a]	m/z value			
	dGPL-I	dGPL-II	dGPL-III	dGPL-IV
$[M+Na]^+$	1312	1309	1309	1306
$[M+H]^+$ after β-elimination of 6dTal	1075	1072	1072	1069
N-Acyl-Phe-(6dTal)Thr-Ala-CO^+	1007	1004	1007	1004
N-Acyl-Phe-ΔThr-Ala-CO^+	792	789	792	789
N-Acyl-Phe-(6dTal)Thr-CO^+	919	916	919	916
N-Acyl-Phe-Thr-CO^+	722	719	722	719
N-Acyl-Phe-ΔThr-CO^+	704	701	704	701
N-Acyl-Phe-CO^+	604	601	604	601
Loss of *N*-acyl-Phe with H transfer to the fragment ion	687	687	684	684

[a] ΔThr indicates loss of the attached 6dTal via β-elimination; ions ending with $-CO^+$ correspond to acylium ions resulting from mass spectrometric cleavage of peptide bonds.

(positive-ion mode) corresponding to an $[M + Na]^+$ molecular ion and a protonated *N*-acylpeptide core fragment ion resulting from losses of the saccharide moieties via β-cleavages (data not shown). The $[M + Na]^+$ molecular ions of dGPL-I, -II, -III, and -IV at m/z 1159, 1173, 1173 and 1187, respectively, are consistent with the chemical analysis and NMR data which localized the differences among the dGPLs to the presence or absence of *O*-methyl groups (14 u) on the 2-position of the terminal 3,4-di-*O*-methyl-α-L-Rha*p* and, possibly, on the hydroxyl group of the *N*-acyl chain. Thus, dGPL-I does not contain any of the additional *O*-methyl groups while dGPL-IV carries both. The former afforded a protonated *N*-acylpeptide core of m/z 817, whereas the latter yielded the corresponding ion at m/z 831. Although both dGPL-II and -III gave molecular ions of identical m/z value, the *N*-acylpeptide ion of GPL-II was present at m/z 831 while that of dGPL-III was seen at m/z 817, suggesting that the single additional *O*-methyl group was localized to a different moiety, viz., the *N*-acyl chain for dGPL-II but the 3,4-di-*O*-methyl-α-L-Rha*p* for dGPL-III.

The full sequences of the dGPLs were afforded by further FABMS analysis of the per-*O*-deuteriomethyl derivatives, and the data are summarized in Table 1. All four dGPLs yielded similar spectra with characteristic fragment ions exhibiting mass shifts depending on the location of the additional *O*-methyl group(s). The spectrum of per-*O*-deuteriomethylated dGPL-IV is shown in Fig. 1 with the assignment of the key fragment ions illustrated schematically. Most importantly, the *N*-terminal fragment ions resulting from the cleavages at each of the three peptide bonds unambiguously defined the peptide sequence, the attachment sites of the monosaccharide moieties, and the molecular weight of the hydroxylated *N*-acyl chain. Thus, for dGPL-IV, the ions at m/z 601, 916, and 1004 together with the ions at m/z 719 (loss of 6-deoxy-Tal from m/z 916 via β-cleavage), 701 and 789 (loss of 6-deoxy-Tal from 916 and 1004, respectively, via β-elimination) unequivocally defined the sequence *N*-acyl-Phe-(6-deoxy-Tal)Thr-Ala-. The attachment of the di- or tri-*O*-methylated Rha residue to the

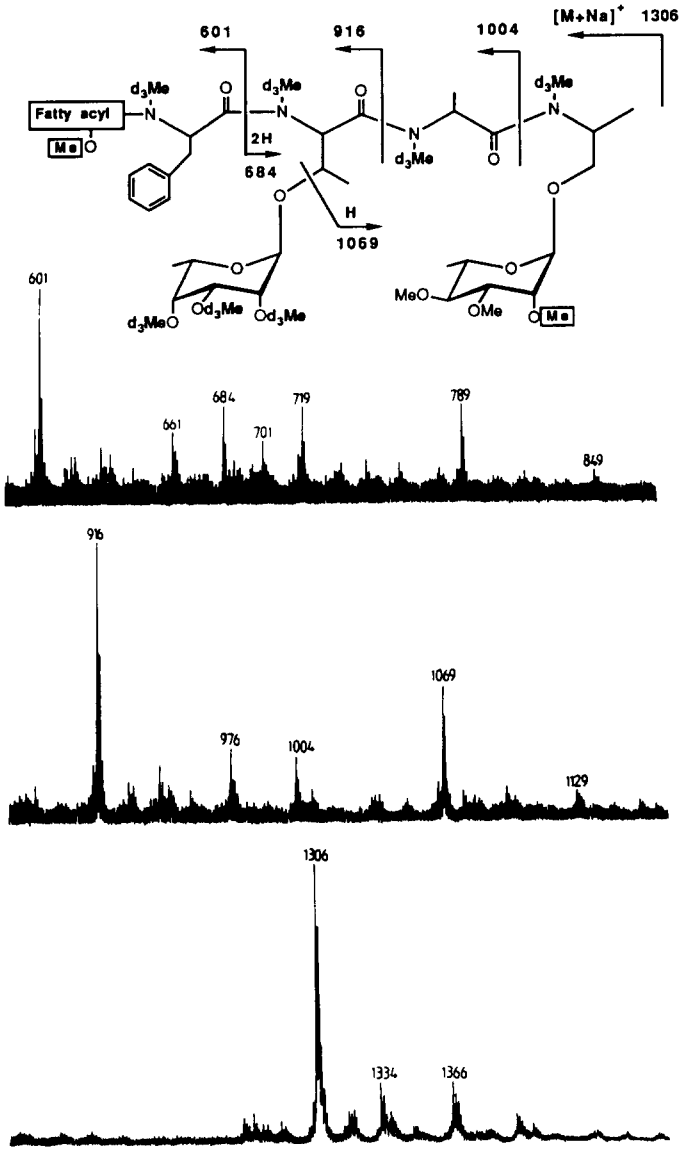

Fig. 1. Positive-ion FAB mass spectrum of per-*O*-deuteriomethylated dGPL-IV from *M. butyricum*. Signal assignment is illustrated schematically. The *O*-methyl groups which may be absent in other GPLs are boxed. dGPL-I lacks both *O*-methyl groups; dGPL-II lacks the *O*-methyl on Rha; dGPL-III lacks the *O*-methyl on the *N*-acyl chain. Mass shifts associated with each of the ions for different dGPLs are tabulated in Table 1. Signals at 60 u higher than those assigned correspond to adducts of unknown nature.

"C-terminal" alaninol was consistent with the mass difference between the molecular ion (*m*/*z* 1306) and the acylium ion (*m*/*z* 1004). Finally, with a full sequence of *N*-acyl-Phe-(6-deoxy-Tal) Thr-Ala-(di/tri-*O*-methyl-Rha)alaninol firmly established, the

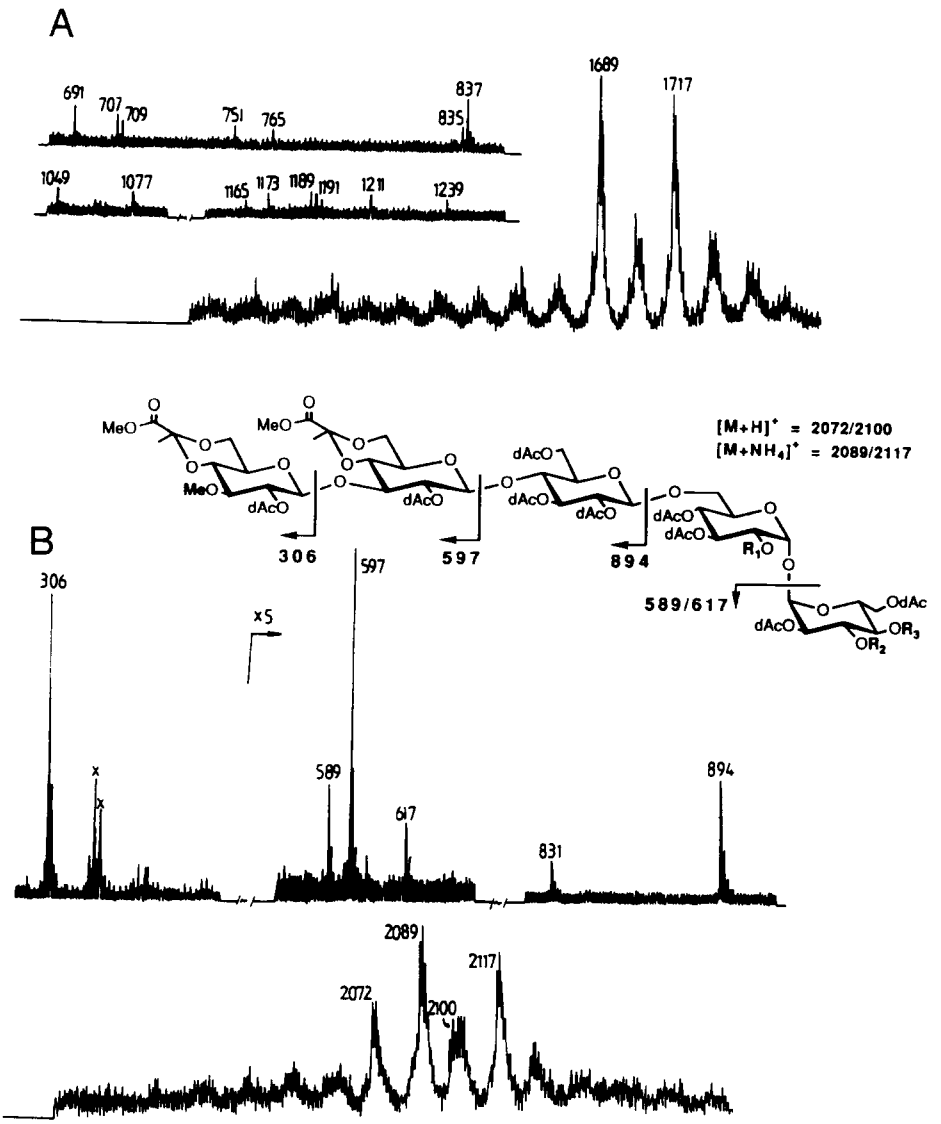

Fig. 2. Positive-ion FAB mass spectra of underivatized (A) and per-*O*-deuterioacetylated (B) LOS from *M. butyricum*. Key fragment ions observed are identical with those afforded by the LOS from *M. smegmatis* mc^{2}11, the assignment of which has been reported [10]. For the underivatized sample (A), the set of glycosidic and ring cleavage ions at *m/z* 691, 707, 709, 751 and 765 correspond to [(MePyr, *O*-methyl)Hex-(MePyr)Hex-Hex]. The analogous set resulting from cleavages at the next Hex residue were observed at *m/z* 837, 1173, 1189 and 1191, corresponding to [(MePyr, *O*-methyl)Hex-(MePyr)Hex-Hex-(2-*O*-C_{22} acyl)Hex], except for the first ion (*m/z* 837), which arises from further eliminating the 2-*O*-C_{22} acyl substituent. In addition, the triacyltrehalose core afforded sodiated ring cleavage ions [(2-*O*-C_{22} acyl)Hex-(3,4-di-*O*-acyl)Hex] at *m/z* 1211/1239 and [Hex-(2-*O*-C_{22} acyl)Hex-(3,4-di-*O*-acyl)Hex] at *m/z* 1049/1077. The 28 mass unit difference observed for these fragment ions and for the molecular ions is attributed to fatty acyl heterogeneity. The key fragment ion signals afforded by the per-*O*-deuterioacetyl derivatives (B) are illustrated schematically. dAc represents the trideuterioacetyl groups incorporated on free hydroxyl groups through derivatization. $R_1 = C_{22}$ fatty acyl; R_2, $R_3 = C_8$ and C_{14}/C_{16} fatty acyl.

hydroxylated *N*-acyl chain length was thus defined as C_{28}, which gives a mass increment of 422 or 436 (*O*-methylated) u.

Our chemical, NMR and FABMS analyses on the LOS from *M. butyricum* indicate that it is structurally identical with the LOS from the phage-resistant $mc^2$11 *M. smegmatis*, the detailed structural data of which we have recently reported [10]. In particular, the key fragment ions afforded by the FAB mass spectra of both the underivatized and per-*O*-deuterioacetyl derivative (Fig. 2) firmly establish the saccharide sequence and the location of the fatty acyl chains of this pyruvylated, acyltrehalose-containing LOS as 4,6-*O*-(1-methoxycarbonylethylidene)-3-*O*-methyl-β-D-Glc*p*-(1 → 3)-4,6-*O*-(1-methoxycarbonylethylidene)-β-D-Glc*p*-(1 → 4)-β-D-Glc*p*-(1 → 6)-2-*O*-acyl-α-D-Glc*p*-(1 ↔)-3,4-di-*O*-acyl-α-D-Glc*p*. The characteristic chemical shifts of the methyl protons of the 4,6-substituted pyruvate ketals in the ^{1}H NMR spectrum were indicative of an *S*-configuration [10]. The single 2-*O*-fatty acyl function was defined as C_{22} (2,4-dimethyl-2-eicosenoate), while the other two fatty acyl chains were defined as C_{14}/C_{16} (tetradecanoate/hexadecanoate) and C_8 (octanoate), based on both FABMS and GC–MS data.

To date, two major families of GPLs have been described in the "atypical" mycobacteria (i.e., non-tuberculosis; usually environmental; many of them the cause of serious opportunistic infections). The best known, the prototypes, are the C-mycosides of the *M. avium* complex which are either apolar, with only one glycosyl residue attached to *allo*-Thr, or polar, with a haptenic oligosaccharide further attached to this single sugar residue which is usually 6-deoxy-L-Tal*p* [2]. Those now identified in *M. butyricum* and previously in *M. peregrinum* [13,14], *M. senegalense* [9,15], and *M. porcinum* [15] all share the same *N*-acyl-D-Phe-(*O*-saccharide)-D-*allo*-Thr-D-Ala-L-alaninyl-*O*-(3,4-di-*O*-methyl-α-L-Rha*p*) core structure as those from *M. avium* but consistently display a single 3-*O*-methyl-α-L-Rha*p* attached to the *allo*-Thr. No haptenic versions of these have yet been found. Instead, structural variants arise through additional substituents (Rha, *O*-methyl, sulfate) on the 3,4-di-*O*-methyl-α-L-Rha attached to alaninol. The serine-containing GPLs of *M. xenopi* constitute the second family, which has very distinctive structural features [3–5]. The reported presence of *N*-methyl-*O*-methylserine in the "C-mycosides" of *M. butyricum* [8] was not further substantiated in our present study. Instead, the GPLs from *M. butyricum* were of the C-mycoside class, similar to those from *M. peregrinum*, *M. senegalense* and *M. porcinum* in that their structural variations reside on the 3,4-*O*-methyl-Rha attached to alaninol, although the presence of 6-deoxy-Tal on *allo*-Thr is a characteristic feature of the apolar C-mycosides from the *M. avium* complex. In addition, it is interesting to note that *M. butyricum* also synthesizes a LOS which is identical with that from a phage-resistant strain ($mc^2$11) of *M. smegmatis* described earlier [10]. The role of these glycolipids in the peculiar life cycle of mycobacteria is still uncertain, perhaps best addressed by genetic transformation of heterologous strains [7].

1. Experimental

M. butyricum *and purification of glycolipids.—M. butyricum* (ATCC 19979 and 559) was grown in a medium containing glycerol, alanine, and salts [16] for 5–7 days at

37°C. Both cells and medium were harvested and the resulting solids extracted twice with $CHCl_3$–CH_3OH (2:1). The dGPLs and LOS were purified from the washed lipid fraction as described previously [2,3,10].

Analytical procedures.—GLC and GLC–MS analyses of the amino acid composition, glycosyl composition, linkages, and enantiomeric configuration, ^1H NMR analysis, and FABMS analyses of native and derivatized samples were all performed as described previously [9,10].

Acknowledgements

This work was supported by Grants AI-18357 and AI-35220 from the National Institute of Allergy and Infectious Diseases, NIH (to P.J.B. and G.S.B.), and a Medical Research Council Programme Grant and a Wellcome Trust Grant 030826 (to H.R.M. and A.D.). K.-H.K. was the recipient of a Wellcome Trust Prize Fellowship (036485/Z/92/Z).

References

[1] P.J. Brennan, in L. Lieve and D. Schlessinger (Eds.), *Microbiology*, American Society for Microbiology, Washington, DC, 1984, pp 203–298.
[2] G.S. Besra and P.J. Brennan, in C. Fenseleau (Ed.), *Mass Spectrometry for the Characterization of Microorganisms*, American Chemical Society, Washington, DC, 1994, pp 203–232.
[3] G.S. Besra, M.R. McNeil, B. Rivoire, K.-H. Khoo, H.R. Morris, A. Dell, and P.J. Brennan, *Biochemistry*, 32 (1993) 347–355.
[4] M. Rivière and G. Puzo, *J. Biol. Chem.*, 266 (1991) 9057–9063.
[5] M. Rivière, S. Augé, J. Vercauteren, E. Wisingerovà, and G. Puzo, *Eur. J. Biochem.*, 214 (1993) 395–403.
[6] C.R. Horsburg and R.M. Selik, *Am. Rev. Respir. Dis.*, 139 (1989) 4–7.
[7] J.T. Belisle, L. Pascopella, J.M. Inamine, P.J. Brennan, and W.R. Jacobs, Jr, *J. Bacteriol.*, 173 (1991) 6991–6997.
[8] E. Vilkas, A. Rojas, and E. Lederer, *C. R. Acad. Sci.*, 261 (1965) 4258–4260.
[9] G.S. Besra, S.S. Gurcha, K.-H. Khoo, H.R. Morris, A. Dell, M.E. Hamid, D.E. Minnikin, M. Goodfellow, and P.J. Brennan, *Zbl. Bakt. Mikrobiol. Hyg.*, 281 (1994) 415–432.
[10] G.S. Besra, K.-H. Khoo, J.T. Belisle, M.R. McNeil, H.R. Morris, A. Dell, and P.J. Brennan, *Carbohydr. Res.*, 251 (1994) 99–114.
[11] G.J. Gerwig, J.P. Kamerling, and J.F.G. Vliegenthart, *Carbohydr. Res.*, 62 (1978) 349–357.
[12] S.L. Mackenzie and D. Tenaschuk, *J. Chromatogr.*, 111 (1975) 413–415.
[13] L.M. López-Marin, M.-A. Lanéelle, D. Promé, M. Daffé, G. Lanéelle, and J.-C. Promé, *Biochemistry*, 30 (1991) 10536–10542.
[14] L.M. López-Marin, M.-A. Lanéelle, D. Promé, G. Lanéelle, J.-C. Promé, and M. Daffé, *Biochemistry*, 31 (1992) 11106–11111.
[15] L.M. López-Marin, M.-A. Lanéelle, D. Promé, and M. Daffé, *Eur. J. Biochem.*, 215 (1993) 859–866.
[16] K. Takayama, H.K. Schnoes, E.L. Armstrong, and R.W. Boyle, *J. Lipid Res.*, 16 (1975) 308–317.