Microbiology (1998), **144**, 589–591 Printed in Great Britain

Inteins in mycobacterial GyrA are a taxonomic character

Peter Sander,[1] Fernando Alcaide,[2] Ingeborg Richter,[1] Klaus Frischkorn,[1] Enrico Tortoli,[3] Burkhard Springer,[1] Amalio Telenti[2] and Erik C. Böttger[1]

Author for correspondence: Peter Sander. Tel: +49 511 532 4361. Fax: +49 511 532 4366.
e-mail: Sander.peter@mh-hannover.de

[1] Institut für Medizinische Mikrobiologie, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, 30625 Hannover, Germany

[2] Département de Medicine Interne, Division des Maladies Infectieuses, CHUV-1011 Lausanne, Switzerland

[3] Bacteriological and Virological Laboratory, Careggi Hospital, 50139 Firenze, Italy

The A subunit of DNA gyrase in mycobacteria is frequently subjected to splicing events as its gene, *gyrA*, harbours an insertion encoding an intein. Investigation of a number of different isolates of *Mycobacterium kansasii*, *Mycobacterium malmoense*, *Mycobacterium marinum*, *Mycobacterium ulcerans* and *Mycobacterium xenopi* demonstrated that the presence of GyrA inteins is not random but a taxonomic character specific for a given taxon at a species or subspecies level.

Keywords: GyrA, intein, mycobacteria, *Mycobacterium kansasii*

INTRODUCTION

The term intein refers to unusual protein sequences that are excised from the precursor protein during maturation. Its coding sequence is always inserted in-frame with a protein-coding sequence (Perler *et al*., 1994). Protein maturation involves excision of the central protein from the precursor molecule and ligation of the N- and C-terminal domains to form the mature protein. Inteins probably possess endonuclease activity since they show homology to endonucleases of eukaryotes which mediate homing of the encoding sequence (Colston & Davis, 1994).

Inteins have so far been described in the yeast vacuolar proton-pump ATPase subunit (*Saccharomyces cerevisiae* and *Candida tropicalis*) and in archaeobacterial DNA polymerases (*Thermococcus litoralis* and *Pyrococcus* spp.) (for a review see Colston & Davis, 1994). Recently, as many as 18 putative inteins have been identified in the genome of *Methanococcus jannaschii*, an archaeon, from which the complete genome sequence has been determined (Bult *et al*., 1996).

In mycobacteria, inteins were first observed in the RecA protein of *Mycobacterium tuberculosis* and *Mycobacterium leprae* (Davis *et al*., 1992). More recently, it was found that inteins may be present in the A subunit of the mycobacterial DNA gyrase (Fsihi *et al*., 1996). The A subunit of the DNA gyrase, GyrA, is well-conserved among bacteria and mediates double-strand breakage and reunion of DNA, while the B subunit is responsible for energy transduction via ATP hydrolysis (for review see Wigley, 1995). The intein coding sequences localize to the same position in *gyrA* in different mycobacterial species, suggesting a specific homing site (Fsihi *et al*., 1996).

Among mycobacteria, the presence of GyrA inteins apparently is not uniform. While in *gyrA* of *M. leprae* intein coding sequences have been constantly observed, inteins have been found only occasionally in the corresponding proteins of *Mycobacterium kansasii*, *Mycobacterium flavescens* and *Mycobacterium gordonae*. Other mycobacterial species were investigated at a single strain level, thus excluding conclusions concerning the spreading of this characteristic feature (Fsihi *et al*., 1996). The finding that some strains of a given species have GyrA inteins, while others have not, prompted us to investigate the nature of this phenomenon.

METHODS

Strains. Several strains, isolated at Medizinische Hochschule Hannover (Germany), Zürich (Switzerland), Berne (Switzerland), Antwerpen (Belgium) and Firenze (Italy), each of *Mycobacterium kansasii* ($n = 23$), *Mycobacterium malmoense* ($n = 8$), *Mycobacterium marinum* ($n = 8$), *Mycobacterium ulcerans* ($n = 4$) and *Mycobacterium xenopi* ($n = 10$) were included in this analysis.

Identification. Isolates were identified by conventional methods as well as by 16S rRNA sequencing (Rogall *et al*., 1990). Strains of *M. kansasii* were also investigated by

0002-1740 © 1998 SGM
The EMBL accession number for the sequence reported in this paper is AJ002066.

hybridization with a commercially available probe for *M. kansasii* (Gen-Probe; Tortoli *et al*., 1994a, b).

Molecular biological methods and protein alignment. PCR amplification of the non-coding 16S–23S rRNA intergenic spacer region was performed with primers 248 (5′-GTG-TGGGTTTCCTTCCTTGG-3′) and 42 (5′-CCACACGGG-TTAACCTCGC-3′) using standard conditions (Rogall *et al*., 1990). Nucleic acid sequencing was done manually using ^{32}P-labelled dCTP and sequenase (USB).

A previously devised PCR strategy using primers H49 (5′-AGGTTGTGCGGCGGGATATTGGT-3′) and H50 (5′-TTCCGCCCGGACCGCAGCCACG-3′) was used to investigate the presence of intein coding sequences in *gyrA* (Fsihi *et al*., 1996). Sequencing of *gyrA* intein coding sequence of *M. malmoense* was performed by PCR-mediated *Taq* cycle sequencing using an ABI373 sequencer. Protein alignment of known GyrA inteins (*M. leprae*, accession no. Z68206; *M. flavescens*, accession no. Z68209; *M. gordonae*, accession no. Z68208; *M. kansasii*, accession no. Z68207; and *M. xenopi*, accession no. U67876) was performed with the program CLUSTAL (PC/Gene, IntelliGenetics, release 6.85) as was the 16S rRNA sequence alignment of these species (Rogall *et al*., 1990).

RESULTS AND DISCUSSION

Mycobacteria with intein coding sequences in *gyrA* yield a PCR product of 1·6 kbp or 0·9 kbp, whereas those with inteinless *gyrA* generate smaller fragments of around 350 bp.

Representative PCR results are presented in Fig. 1. The different isolates of *M. malmoense*, *M. marinum*, *M. ulcerans* and *M. xenopi* investigated showed a constant pattern: (i) isolates of *M. malmoense* were characterized by a gene fragment of 1.6 kbp, indicative of a GyrA intein; (ii) isolates of *M. marinum* and *M. ulcerans* showed an amplified gene fragment of 350 bp characteristic for GyrA lacking an intein; (iii) isolates of *M. xenopi* presented a 0.9 kbp fragment corresponding to a shorter intein coding sequence in *gyrA*, unique to this species. In contrast, isolates of *M. kansasii* were heterogeneous, producing either a 350 bp or a 1·6 kbp amplification fragment: 10 isolates were inteinless, while 13 isolates, including the *M. kansasii* type strain ATCC 12478, were characterized by a 1·6 kbp PCR product.

M. malmoense, *M. marinum*, *M. ulcerans* and *M. xenopi* are defined by a 16S rRNA sequence which is identical for different isolates of a given species, while two distinct 16S rRNA gene sequences have been reported for *M. kansasii* (Ross *et al*., 1992). Previously, the existence of two genetically distinct subspecies of *M. kansasii* has been suggested. These two subtypes can be differentiated by 16S rRNA gene sequence determination (Ross *et al*., 1992), by hybridization to a probe derived from a repetitive DNA element (Yang *et al*., 1993b) or by hybridization to a commercially available probe for identification of *M. kansasii* (Gen-Probe; Yang *et al*., 1993a, b; Tortoli *et al*., 1994a). For the purpose of this study, these two groups of *M. kansasii* defined by 16S rRNA and probe analysis are referred to as type I (Gen-Probe positive) and type II (Gen-Probe negative). The two different 16S rRNA sequences found within our collection of *M. kansasii* isolates showed a complete correlation with the results of the probe assay.

Sequence analysis of the non-coding 16S–23S rRNA intergenic spacer region showed identical nucleic acid sequences for each of the different isolates of *M. malmoense*, *M. marinum* and *M. ulcerans*, while sequence heterogeneity was observed for the different isolates of *M. kansasii* (data not shown). This sequence heterogeneity matched the categorization by 16S rRNA and genetic probe analysis. Thus, in contrast to *M.*

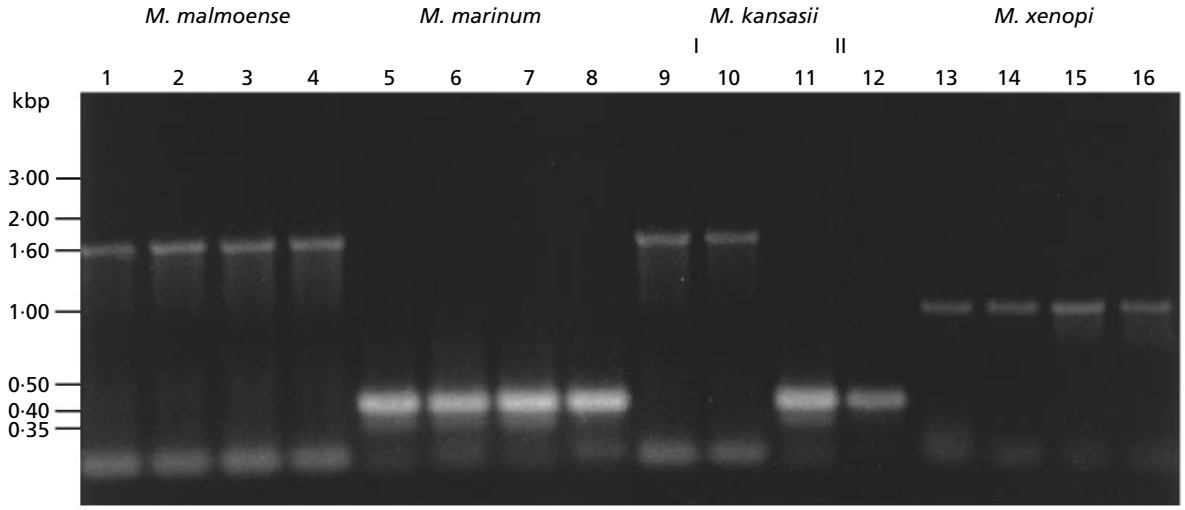

Fig. 1. PCR amplification of a *gyrA* gene fragment from different mycobacteria. Mycobacteria with intein coding sequences in *gyrA* yield a PCR product of 1·6 kbp or 0·9 kbp, whereas those with inteinless GyrA generate a smaller fragment of around 350 bp. Lanes: 1–4, *M. malmoense*; 5–8, *M. marinum*; 9 and 10, *M. kansasii* type I; 11 and 12, *M. kansasii* type II; 13–16, *M. xenopi*.

malmoense, *M. marinum* and *M. ulcerans*, which appear as genetically homogeneous taxons, *M. kansasii* genetically is more diverse and consists of two genetically distinct subspecies.

Analysis of the PCR amplification data (see Fig. 1) demonstrated that the presence of an intein coding sequence in *gyrA* was invariably associated with *M. kansasii* type I (13 of 13 investigated), whereas none of the type II *M. kansasii* isolates showed the presence of an intein coding sequence (10 of 10 investigated).

Sequence determination of PCR products obtained with primers H49 and H50 (Fsihi *et al.*, 1996) and *M. malmoense* genomic DNA as template revealed an insertion following codon 130 of *gyrA* (Perler *et al.*, 1997); the insertion comprises 1260 bp, encoding a putative intein of 420 aa. Comparison with *gyrA* genes of other mycobacterial species indicates a conservation of size (420 aa in *M. leprae*, *M. kansasii* type I and *M. gordonae* and 421 aa in *M. flavescens*), the N-terminal amino acid (Cys) and the C-terminal splice junction (His Asn/Thr). These GyrA inteins are much longer than the *M. xenopi* GyrA intein (198 aa), which lacks homing endonuclease activity and has undergone a complex series of recombination events (Telenti *et al.*, 1997).

Phylogenetic trees based on 16S rRNA sequences indicate a closer relationship of *M. malmoense*, *M. gordonae* and *M. kansasii* to one another than to *M. flavescens* or *M. leprae*. A similar relationship was found when the GyrA intein sequences were compared (data not shown).

From our analysis, we draw the conclusion that the presence of a GyrA intein is not random but is a taxonomic character. This character defines a mycobacterial species either at a species or at a subspecies level depending on the degree of genetic homogeneity within a given taxon. Genetically homogeneous species such as *M. malmoense*, *M. marinum*, *M. ulcerans*, *M. tuberculosis* and *M. leprae* uniformly are characterized by the presence or absence of a GyrA intein. *M. xenopi* is characterized by its unique short *gyrA* element. In genetically more heterogeneous species, such as *M. kansasii* – and probably *M. flavescens* and *M. gordonae* (Fsihi *et al.*, 1996) – the presence or absence of an intein coding sequence in *gyrA* corresponds to a subspecies characteristic. Our findings do not imply that the intein itself – although a valid taxonomic marker – is useful for defining phylogenetic relationships as the underlying mechanism of acquisition or loss of that marker remains to be determined.

ACKNOWLEDGEMENTS

We are indebted to Francoise Portaels and Gaby Pfyffer for providing strains. We thank Kerstin Teschner for excellent technical assistance.

This work was supported in part by grants from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technik (Verbund Mykobakterielle Infektionen). E. C. B. is a Hermann- und Lilly-Schilling Professor of the Stifterverband für die Deutsche Wissenschaft. A. T. is supported by the Swiss National Science Foundation.

REFERENCES

Bult, J. C., White, O., Olsen, G. J. & 37 other authors (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073.

Colston, M. J. & Davis, E. O. (1994). The ins and outs of protein splicing elements. *Mol Microbiol* **12**, 359–363.

Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, M. J. & Sedgwick, S. G. (1992). Protein splicing in the maturation of *M. tuberculosis* RecA protein: a mechanism for tolerating a novel class of intervening sequence. *Cell* **71**, 201–210.

Fsihi, H., Vincent, V. & Cole, S. T. (1996). Homing events in the *gyrA* gene of some mycobacteria. *Proc Natl Acad Sci USA* **93**, 3410–3415.

Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J. & Belfort, M. (1994). Protein splicing elements: inteins and exteins – a definition of terms and recommended nomenclature. *Nucleic Acids Res* **22**, 1125–1127.

Perler, F. B., Olsen, G. J. & Adam, E. (1997). Compilation and analysis of intein sequences. *Nucleic Acids Res* **25**, 1087–1093.

Rogall, T., Flohr, T. & Böttger, E. C. (1990). Differentiation of *Mycobacterium* species by direct sequencing of amplified DNA. *J Gen Microbiol* **136**, 1915–1920.

Ross, B. C., Jackson, K., Yang, M., Sievers, A. & Dwyer, B. (1992). Identification of a genetically distinct subspecies of *Mycobacterium kansasii*. *J Clin Microbiol* **30**, 2930–2933.

Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W. R. & Perler, F. B. (1997). The *Mycobacterium xenopi GyrA* protein splicing element: characterization of a minimal intein. *J Bacteriol* **179**, 6378–6382.

Tortoli, E., Simonetti, M. T., Lacchini, C., Penati, V., Piersimoni, C. & Morbiducci, V. (1994a). Evaluation of a commercial DNA probe assay for the identification of *Mycobacterium kansasii*. *Eur J Clin Microbiol Infect Dis* **13**, 264–267.

Tortoli, E., Simonetti, M. T., Lacchini, C., Penati, V. & Urbano, P. (1994b). Tentative evidence of AIDS-associated biotype of *Mycobacterium kansasii*. *J Clin Microbiol* **32**, 1779–1782.

Wigley, D. B. (1995). Structure and mechanism of DNA topoisomerases. *Annu Rev Biophys Biomol Struct* **24**, 185–208.

Yang, M., Ross, B. C. & Dwyer, B. (1993a). Identification of an insertion sequence-like element in a subspecies of *Mycobacterium kansasii*. *J Clin Microbiol* **31**, 2074–2079.

Yang, M., Ross, B. C. & Dwyer, B. (1993b). Isolation of a DNA probe for identification of *Mycobacterium kansasii*, including the genetic subgroup. *J Clin Microbiol* **31**, 2769–2772.

Received 1 April 1997; revised 23 September 1997; accepted 1 October 1997.