FEBS Letters 394 (1996) 206–212 FEBS 17559

Identification of a second *Mycobacterium tuberculosis* gene cluster encoding proteins of an ABC phosphate transporter

Martine Braibant[a], Philippe Lefèvre[a], Lucas de Wit[a], Josette Ooms[a], Priska Peirs[a], Kris Huygen[a], Rudy Wattiez[b], Jean Content[a,*]

[a] *Institut Pasteur, Department of Virology, rue Engeland 642, B-1180 Brussels, Belgium*
[b] *Université de Mons Hainaut, Faculté des Sciences, Service de Chimie Biologique, av. Maistriau 21, B-7000 Mons, Belgium*

Received 18 June 1996; revised version received 21 August 1996

Abstract Following the identification of a *M. tuberculosis* phosphate transporter belonging to the superfamily of ABC transporters, we report on the cloning and sequencing of two additional genes, called *pstS-3* and *pstC-2*, encoding proteins homologous to PstS and PstC of *Escherichia coli*, respectively. Together with the previously isolated *M. tuberculosis* gene similar to the *E. coli pstA*, these are included in a cluster encoding a second putative phosphate transport system. We demonstrate that *pstS-3* encodes the previously described Ag 88, a 40 kDa *M. bovis* BCG culture filtrate antigen (immunodominant in H-2^b haplotype type mice). Finally, a signature motif identifying integral transmembrane proteins of prokaryotic phosphate binding-dependent permeases is proposed.

Key words: ATP-binding-cassette transporter; Phosphate transport; PstS; PstC; Ag 88

1. Introduction

Tuberculosis is still a major health problem worldwide. Each year eight million people develop the disease and it is estimated that tuberculosis currently accounts for approximately three million deaths annually. Strains of *M. tuberculosis* resistant to existing drugs are emerging, leading to the urgent need for the development of new anti-mycobacterial agents [1,2].

In Gram-negative bacteria, many ligand-binding-protein-dependent transport systems for sugars, amino acids, peptides, ions and vitamins have been identified. These transport systems are multisubunit permeases composed of a soluble substrate-binding protein and a membrane-bound complex containing two to four proteins. They are members of a superfamily of membrane transporters called ABC (ATP-binding cassette) transporters, which includes eukaryotic proteins such as the multidrug resistance (MDR) protein, the cystic fibrosis transmembrane regulator (CFTR) [3] and TAP transporters required for the presentation of endogenous antigens by HLA class I molecules [4]. In Gram-positive bacteria, several high affinity transport systems with an analogous organization have been identified [5,6]. These organisms, which have no outer membrane, maintain the extra-cytoplasmic protein in the vicinity of the cytoplasmic membrane by means of an NH_2 terminal lipo-amino acid anchor.

The *Escherichia coli* phosphate-specific transport (Pst) system is typical of this class of permeases and comprises four distinct subunits encoded by the *pstS*, *pstA*, *pstB* and *pstC* genes. The PstS protein is a phosphate-binding protein located in the periplasmic space. The PstA and PstC proteins are hydrophobic and are considered to form the transmembrane channel of the Pst system. The ATP binding PstB protein does not possess the hydrophobic characteristics of a membrane protein, and probably interacts with PstA-PstC at the cytoplasmic side of the membrane [7–9].

A *M. tuberculosis* lipoprotein (named Ag 78, pab, 38 kDa protein) equivalent to *E. coli* PstS phosphate-binding component was previously described [10–12]. This protein has been expressed in *E. coli* and its affinity for the phosphate anion has been determined [13]. Recently, we demonstrated that the gene encoding this protein is part of an operon of four genes (*pstB*, *pstS-1**, *pstC-1* and *pstA-2*) and that phosphate uptake by whole *M. bovis* BCG is inhibited by a monoclonal antibody directed against the PstS-1 component [14]. Since the cloning and sequencing of a *M. tuberculosis* gene encoding the first homolog of *E. coli* PstA was described earlier [15] and more recently, a second PstS gene equivalent, called *pstS-2*, was identified in our laboratory (Lefèvre et al., submitted), we examined genomic regions surrounding these two genes attempting to identify other Pst subunits.

Here we report that the first isolated *M. tuberculosis* gene, *pstA-1*, similar to *E. coli pstA* [15] is also included in a cluster of three genes (*pstS-3*, *pstC-2* and *pstA-1*) encoding a second putative phosphate permease. Furthermore, we demonstrate unambiguously that the *pstS-3* gene encodes the previously described Ag 88, a 40-kDa antigen of *M. bovis* BCG culture filtrate [16,17], mistakenly considered to be encoded by *pstA-1* in a previous report [15]. We also define a signature motif for integral cytoplasmic membrane proteins of prokaryotic phosphate binding protein-dependent permeases.

2. Materials and methods

2.1. Screening of the λgt11 *M. tuberculosis* recombinant DNA library by hybridization

A λgt11 recombinant library constructed from genomic DNA of *M. tuberculosis* (Erdman strain) was obtained from R.A. Young [18]. Screening was performed as described before by plaque hybridization [19] with a 346 bp *Pst*I-*Pvu*II fragment of the previously isolated A1 clone used to sequence the *pstA-1* gene [15].

2.2. Subcloning and sequencing

Specific fragments were subcloned in pBluescript II SK+. Sequence analysis was performed on both strands using Sanger's technique with Taq DNA polymerase (Promega) and 7-deaza-dGTP instead of dGTP and by Texas Red dye-primer cycle sequencing on an automatic se-

*Corresponding author. Fax: +32 2 3733279.
e-mail: jcontent@ben.vub.ac.be.

PII S0014-5793(96)00953-2

quencer using Thermosequenase Core Sequencing Kit with 7-deaza-dGTP (Vistra DNA Systems, Amersham).

2.3. Nucleotide sequences accession numbers

The DNA sequences of *pstS-3* and *pstC-2* genes have been deposited in the European Molecular Biology Laboratory data base under accession numbers Z48057 and Z47983 respectively.

2.4. Affinity purification of Ag 88

Monoclonal antibody 2C1-5 directed against Ag 88 [16] was purified from 500-ml hybridoma culture supernatants using an AvidChrom Protein A Antibody purification Kit (Sigma), then coupled to an Hydrazide AvidChrom Cartridge (Sigma). Culture filtrate antigens were obtained as described previously [17] by ammonium sulfate precipitation of the culture supernatant of *M. bovis* BCG GL2 grown for 2 weeks as a surface pellicle on Sauton culture medium. About 7 mg culture filtrate antigens were applied to the cartridge. Bound material was eluted by lowering the pH to 2 with 50 mM glycine-HCl buffer. The eluted antigen was immediately brought to neutral pH and analysed by Western blot with mAb 2C1-5 using ProtoBlot Western Blot AP Systems (Promega). The purity of the Ag 88 protein was examined on Coomassie blue-stained SDS-polyacrylamide gel.

2.5. Amino acid microsequence analysis

Affinity-purified Ag 88 was separated by SDS-polyacrylamide gel electrophoresis (PAGE) as described by Laemmli [20] and electroblotted onto PVDF membranes by the method of Matsudaira [21]. Prior to in situ CNBr cleavage, the blot was stained with Coomassie blue. In situ CNBr cleavage of the protein was carried out according to the method of Stone and Williams [22]. The mixture of CNBr peptides was eluted from PVDF by the method of Matsudaira [21]. The eluted peptides were separated by SDS-PAGE and electroblotted onto PVDF membranes as described above. Amino acid microsequence analysis was performed by automated Edman degradation of 1–10 pmol of peptides on a Beckman LF3400 protein sequencer (Beckman Instruments, Inc). All samples were sequenced using standard Beckman sequencer procedure. All sequencing reagents were from Beckman Instruments, Inc.

2.6. Computer analysis

Computer-aided analysis of the nucleic acid and deduced amino acid sequences were performed with the DNA Strider program [23] and the Genetics Computer Group program [24] of the Belgian EMBnet Node (network facility). Multiple sequence alignments were carried out using the GCG program PILEUP, which generates optimal alignments. Searching databases with a signature was carried out using the FINDPATTERNS program. Candidate signatures were run through the PROT database which contained the whole SwissProt database (release 33+ updated), non-redundant PIR database (release 47) and non-redundant GenPept database (release 94+ updated).

3. Results

3.1. Nucleotide sequence of the operon

We previously suggested that the N-terminal region of *M. tuberculosis* PstA-1 contained a small sequence homologous to the carboxy-terminal region of PstC [15]. Since, in *E. coli*, *pstS*, *pstC*, *pstA* and *pstB* genes, coding for the four subunits of the phosphate permease, are located within a single operon, we sequenced the 1000-bp *Eco*RI fragment of the λgt11 A1 clone in order to reexamine the *pstC-pstA* junction (Fig. 1). This enabled us to delineate a full-size *pstC* gene equivalent immediately adjacent to a *pstA* gene homologue which now starts at codon GTG (position 2177 to 2179, Fig. 2) and overlaps the termination codon TGA of *pstC-2*. The deduced amino acid sequence of *pstC-2* gene corresponds to a protein of 323 amino acids (M_w 34.136 kDa) highly similar to *E. coli* PstC (34.3% identity in a 297 aa overlap and 76.1% similarity when conservative replacements are considered). For other PstC homologues from *Haemophilus influenzae* [25], *Synechocystis* sp. [26], *Bacillus subtilis* (unpublished data, accession number P46339) and *Mycobacterium intracellulare* (unpublished data, accession number X95538), identity varies from 27.9% to 37.4%. The highest identity score was obtained with an open reading frame named 'phoW1' from *M. leprae* cosmid B650 which showed 82.4% identity in a 319 aa overlap and 96.2% similarity (Fig. 1). By comparing the two *M. tuberculosis* proteins, PstC-1 [14] and PstC-2, 28.8% identity in a 313 aa overlap and 73.5% similarity was found. The corresponding DNA region shows 55.8% identity in a 722 nucleotides overlap.

When the hydropathy profiles [27] and predictions of the membrane spanning regions [28] of the two mycobacterial PstC-1 and PstC-2 proteins and the *E. coli* PstC protein are compared, the structural homology of the three proteins is also apparent, suggestive of the existence of six hydrophobic membrane-spanning regions (Fig. 2). The signature of an inner membrane component of binding-protein-dependent transport system was found in this mycobacterial PstC-2 protein between helix IV and V, a region homologous to its location in the *E. coli* PstC protein and in other integral inner-membrane proteins (Fig. 2) [29]. Positively charged residues are found in excess at the cytoplasmic side and could

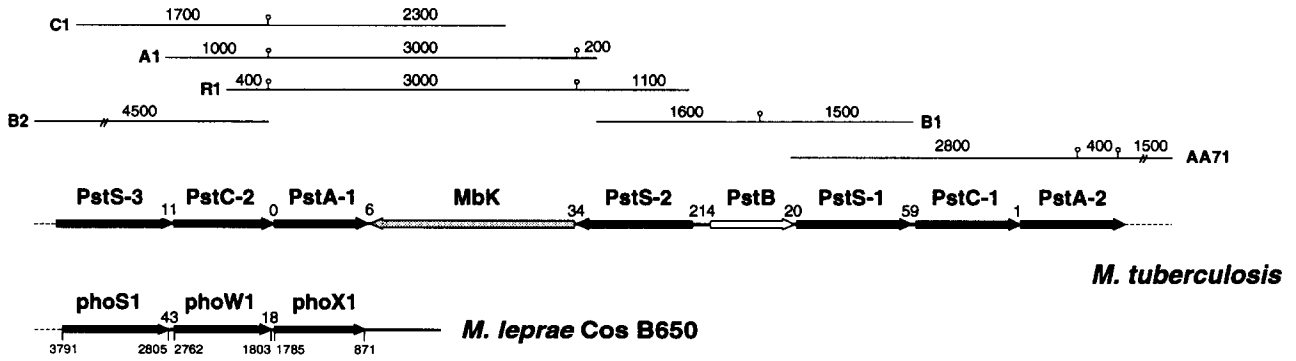

Fig. 1. Schematic view of the genes coding for subunits of three putative phosphate permeases of *M. tuberculosis*. Genes are organized in three potential operons: *pstS-3.pstC-2.pstA-1* (this article), *pstS-2*.MbK (Peirs et al., submitted; Lefevre et al., submitted) and *pstB.pstS-1.pstC-1.pstA-2* [14]. The gene order in the *E. coli* pst operon is *pstS*, *pstC*, *pstA* and *pstB*. The intergenic distances are indicated. C1, A1 [15] (and this article), R1 (Lefèvre et al., submitted), B2 (this article), B1 [14] and AA71 [10] are the different inserts from the λgt11 *M. tuberculosis* library. The *M. leprae* open reading frames 'phoS1', 'phoW1' and 'phoX1' found in cosmid B650 are similar to the three proteins of the operon described in this paper. ♀ represents *Eco*RI restriction sites.

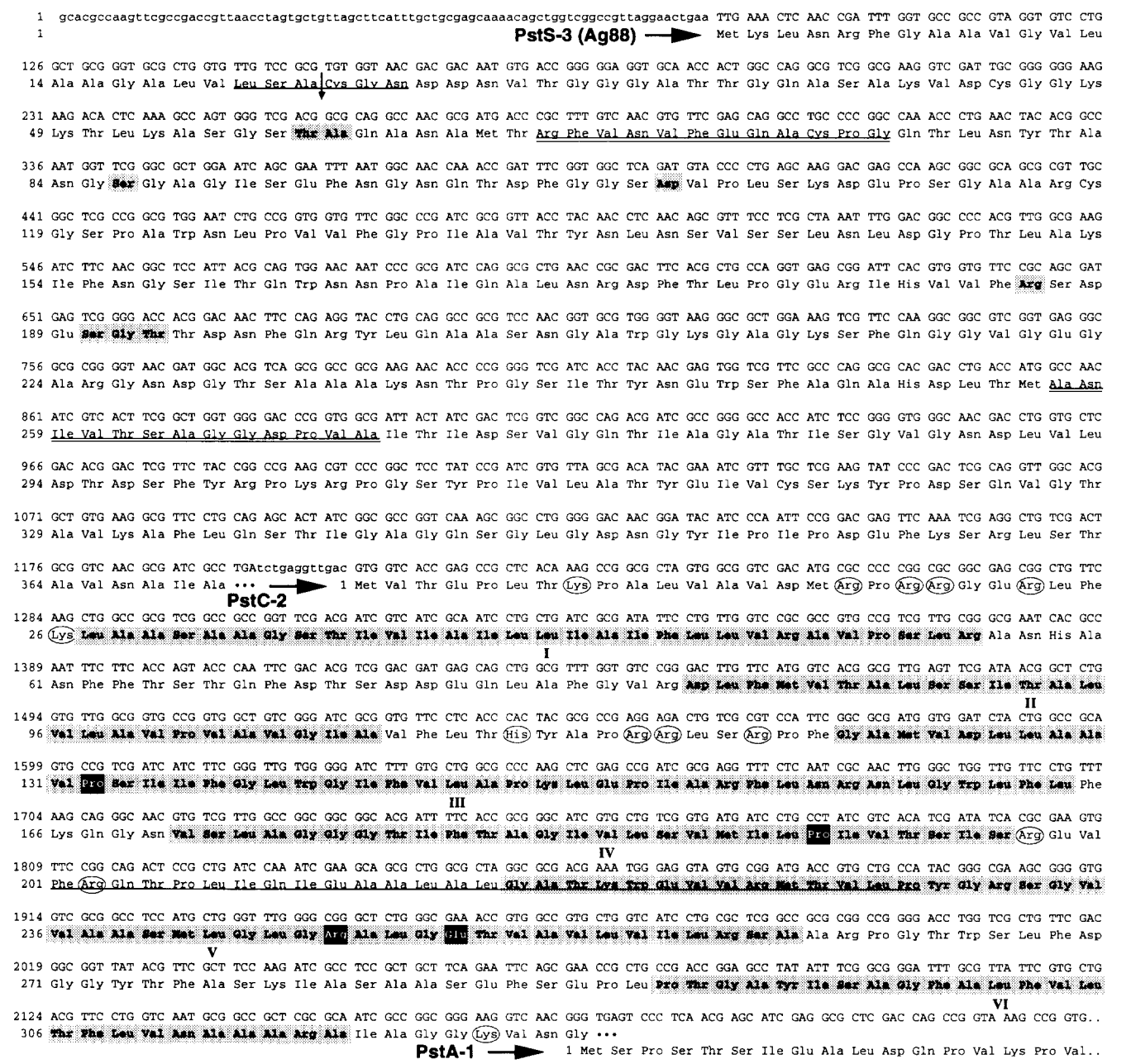

Fig. 2. Nucleotide sequence of *pstS-3, pstC-2* and *pstA-1* genes and deduced amino acid sequences. Lower-case letters represent non-coding DNA regions. In PstS-3, the lipoprotein consensus sequence is underlined, an arrow indicating the cleavage site. The two short peptidic sequences obtained by amino acid sequencing of two CNBr fragments are doubly underlined. Grey boxes represent the eight amino acids probably involved in hydrogen bonding with the phosphate [34]. The signature of the inner membrane components of binding-protein-dependent transport systems, found in PstC-2, is underlined [29]. The six putative α-helices of PstC-2 are represented in grey boxes numbered from I to VI. Positively charged residues found at the putative cytoplasmic side are circled. The conserved essential charged amino acid and proline residues found in the third, fourth and fifth α-helices of PstC-2 are in black boxes [31,41]. The *pstC-2.pstA-1* border represents an extreme situation where the TGA stop codon of *pstC-2* overlaps the GTG start codon of *pstA-1*. The PstA-1 sequence, not presented here completely, refers to the previously published sequence (X75297 accession number).

play a role in orienting the PstC proteins within the membrane (Fig. 2) [30].

In *E. coli*, two charged amino acid residues, Arg-237 and Glu-241, required for phosphate uptake have been identified in the fifth transmembrane α-helix of PstC by directed mutagenesis [31]. Another essential feature of the hydrophobic PstC subunit is a pair of proline residues, Pro-123 and Pro-183, in the putative transmembrane helices three and four. When these prolines are replaced by alanine, the permease appears to be a permanently open transport system. Interestingly, all four essential residues are conserved in the *M. tuberculosis* PstC-2 subunit (Fig. 2).

In *E. coli*, the *pstC* gene is preceded by the *pstS* gene encoding the phosphate binding component. Following a genomic walking approach using a 346 bp *Pst*I-*Pvu*II fragment of the clone A1, recombinant phage B2 was isolated (Fig. 1).

Its digestion with *Eco*RI revealed an insert of 4500 bp which was subcloned into the *Eco*RI site of pBluescript II SK+. Clone B2 was sequenced 1350 nucleotides beyond the A1 extremity. An open reading frame of 369 amino acids was revealed with an expected molecular weight of 37.731 kDa. This sequence is highly similar to *E. coli* PstS (29% identity, 81% similarity) and *M. tuberculosis* PstS-1 (28% identity, 73% similarity) and PstS-2 (63% identity, 94% similarity) (Lefèvre et al., submitted). For other bacterial PstS homologues from *Synechocystis* sp. [26], *Xanthomonas oryzae* [32], *Haemophilus influenzae* [25], *Mycobacterium intracellulare* (unpublished data, accession number X95538), *Synechococcus* sp. [33], identity varies from 24.9 to 31.8%. Again, the highest score was obtained with a *M. leprae* open reading frame named 'phoS1' from cosmid B650 showing 78.7% identity in 328 aa overlap and 94.6% similarity (Fig. 1). This protein was therefore named PstS-3. As for PstS-1 and other ABC transporters in Gram-positive bacteria, this putative ligand-binding subunit is probably acylated and anchored within the cell membrane [6,12]. Indeed, its amino acid sequence predicts a typical 22 aa long hydrophobic signal sequence separated from the mature protein by a lipoprotein consensus sequence (L-*Y*-*Z*-cleavage site-C-y-z; where *Y* is A, S, V, Q, T; *Z* is G or A; y is S, G, A, N, Q, D; z is S, A, N, Q). The predicted molecular weight of the anchored form of PstS-3 is 35.621 kDa.

Recently, the X-ray diffraction structure of the liganded form of *E. coli* PstS has shown the eight amino acids directly involved in hydrogen bonding with the phosphate (Thr-10, Phe-11, Ser-38, Asp-56, Arg-135, Ser-139, Gly-140 and Thr-141) [34]. Interestingly, all are conserved in PstS-3 (Thr-57, Ala-58, Ser-86, Asp-104, Arg-186, Ser-190, Gly-191, Thr-192) except for Phe-11 which is replaced by an Ala residue, another non-polar amino acid (Fig. 2).

Downstream the *M. tuberculosis pstA-1* gene, no homology to *pstB* could be found. However, by sequencing the 3000-bp *Eco*RI fragment of the λgt11 A1 clone, we discovered a gene coding for MbK, a serine-threonine protein kinase (Peirs et al., submitted) (Fig. 1).

3.2. *Ag 88 (PstS-3) is probably a phosphate binding component*

Ag 88, a *M. bovis* BCG culture filtrate antigen of 40 kDa, highly immunodominant in H-2^b haplotype mice [16,17], has been suggested before as homologous to the *E. coli* membrane-bound PstA subunit [15]. Following the identification of its correct start codon, the predicted amino acid sequence of *pstA-1* gene corresponds now to a protein of 305 amino acids with an expected molecular mass of 32.907 kDa. Because of the discrepancy in molecular weight and since it was difficult to understand how such an hydrophobic protein could be found in soluble form in the culture filtrate, we decided to reinvestigate the nature of the *M. bovis* BCG culture filtrate 40 kDa antigen, attempting to purify it from culture filtrate by using the monoclonal antibody 2C1-5 that had been used for cloning of *M. tuberculosis pstA-1* [15]. The results are shown in Fig. 3A,B. A triple band with predicted M_w around 40 kDa could be purified and recognized by the monoclonal antibody 2C1-5. It may correspond to different processed forms of the protein. The two other bands observed at 25 and 55 kDa are light and heavy chains of monoclonal antibodies which broke off the cartridge during elution. Staining of these two bands in Western blot is probably due to recognition of the antibody chains by the secondary anti-mouse IgG antibody. An attempt to sequence the amino terminal end of Ag 88 was unsuccessful. The antigen was therefore cleaved in situ with CNBr, generating two peptides of 22 kDa and 11 kDa that were microsequenced (Fig. 3C). Interestingly, the two amino acid sequences obtained were present in the deduced amino acid sequence of the *pstS-3* gene, suggesting that Ag 88 is not a *M. tuberculosis* homolog of *E. coli* PstA subunit but rather an homolog of the secreted phosphate binding component PstS. Moreover, these amino acid sequences are located just after Met residues and the predicted size of the two peptides generated by CNBr cleavage of PstS-3 is 20.257 kDa and 11.557 kDa, which corresponds to the size of the two Ag 88 CNBr peptides. The only discordance between the sequenced 22 kDa Ag 88 peptide and the 20.257 kDa PstS-3 predicted peptide was Ser instead of Thr. This

Table 1
List of conserved motifs in hydrophobic membrane proteins of phosphate permeases and identification of a signature.

Proteins	Bacteria	Sequence	Reference
PstA-1	*Mycobacterium tuberculosis*	LRLV**P**DEL**RE**A**S**Y**ALG**V**P**KWKT**I**VR**I**A**A**P	[15]
PstA-2	*Mycobacterium tuberculosis*	LAQV**P**TSY**RE**A**AE**A**LG**L**P**AGWA**L**RK**I**V**L**K	[14]
PstC-1	*Mycobacterium tuberculosis*	FRQV**P**VLP**RE**G**A**I**ALG**M**S**NWEC**V**RR**V**TL**P**	[14]
PstC-2	*Mycobacterium tuberculosis*	FRQT**P**LIQ**IE**A**A**L**ALG**A**T**KWEV**V**RM**T**VL**P**	This study
PstA	*Escherichia coli*	LKLV**P**YSL**RE**A**AY**A**LG**T**P**KWKM**I**SA**I**TL**K**	[7]
PstC	*Escherichia coli*	FEQT**P**VMM**KE**S**AY**G**IG**C**T**TWEV**I**WR**I**VL**P**	[7]
PstA	*Haemophilus influenzae*	LLLV**P**NNL**RE**A**AA**A**LG**C**S**QWQV**I**MM**I**C**Y**R**	[25]
PstC	*Haemophilus influenzae*	FSIV**P**PML**KE**G**AY**G**LG**A**T**TWEV**V**RQ**V**IV**P**	[25]
PstA	*Synechocystis* sp.	LQIV**P**QDI**RW**A**AL**G**VG**A**YKYQ**T**V**L**F**V**VL**P**	[26]
PstC	*Synechocystis* sp.	LASL**P**PEL**RQ**A**S**L**GLG**A**T**RWETI**F**RV**L**I**P**	[26]
PstA	*Pseudomonas aeroginosa*	LARI**P**RAV**RE**G**S**L**ALG**A**T**KAET**L**WK**I**VLP	[42]
orf73 (PstA homolog)	*Bacillus subtilis*	IRSV**P**KDL**KE**A**S**L**ALG**V**S**RWHT**V**KT**V**LI**P**	P46340
orf72 (PstC homolog)	*Bacillus subtilis*	MASL**P**KSL**RE**G**SY**A**LG**A**T**RWQTI**R**KV**L**VL**P**	P46339
phoX1 (PstA homolog)	*Mycobacterium leprae*	LRLV**P**DEL**RE**A**C**Y**ALG**I**P**KWKT**I**VR**I**VF**P**	U15184
phoW1 (PstC homolog)	*Mycobacterium leprae*	FRHT**P**LIQ**IE**A**A**Q**ALG**A**T**KWEV**V**QM**T**VL**P**	U15184
PstC1	*Mycobacterium intracellu lare*	IRQV**P**VLP**RE**G**A**V**ALG**M**T**DWEC**A**RR**V**TL**P**	X95538
AG88 homolog (PstA homolog)	*Mycoplasma genitalium*	LNNV**S**WDL**RI**S**A**F**ALG**I**S**KREVI**F**KI**V**LP	[40]
Signature		PXXX**R**X**AA**X**ALG**XT**XXXX**I**XX**I**XX**P** **S K GS GI S V V K** **I SC V P L T R** **Y A**	

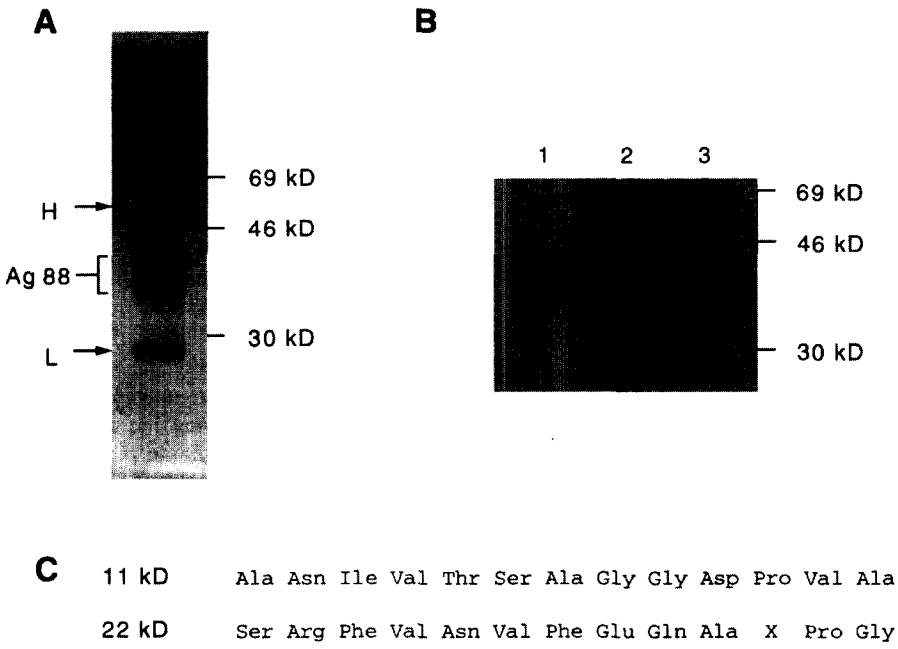

Fig. 3. A: Coomassie-staining following separation on a 15% polyacrylamide gel of Ag 88 purified from *M. bovis* BCG by affinity chromatography with mAb 2C1-5. B: Immunoblotting analysis, with monoclonal antibody 2C1-5 of *M. bovis* BCG culture filtrate (1), *pstS-3* expressed as recombinant protein (2) and Ag 88 purified from culture filtrate (3). C: Microsequence of the 22 kDa and 11 kDa peptides generated by CNBr cleavage of the affinity purified Ag 88.

may be due to a strain difference between *M. bovis* BCG (from which Ag 88 was purified) and *M. tuberculosis* (from which λgt11 recombinant library was derived). Recently, the *M. tuberculosis pstS-3* gene has been expressed in *E. coli* in frame with a 6xHis encoding headpiece and the recombinant protein was purified by metal affinity chromatography on nickel ions (Lefèvre et al., submitted). This recombinant protein is easily detected in Western blot with the monoclonal antibody 2C1-5 directed against Ag 88, confirming that PstS-3 is indeed Ag 88 (Fig. 3B).

Finally, we reexamined the previously isolated C1 λgt11 clone which had been selected with the two monoclonal antibodies, 2C1-5 and 2F8-3, directed against Ag 88 [15]. Its digestion with *Eco*RI had revealed an insert of 2.3 kb containing all the *pstA-1* gene and the 3′ extremity of *pstC-2* gene but not *pstS-3* gene (Fig. 1). Sequencing of the λgt11 C1 border and restriction analysis (*Kpn*I-*Sst*I) demonstrated that a cryptic *Eco*RI had obscured the existence of an 1.7 kb *Eco*RI fragment which contains *pstC-2* and the major part of *pstS-3* (278 of 369 amino acids). Therefore, the fusion protein of 150–155 kDa expressed in the *E. coli* lysogens [15] and recognized by monoclonal antibodies 2F8-3 and 2C1-5 was certainly a fusion between β-galactosidase and 75% of PstS-3.

3.3. Characterization of a signature for integral cytoplasmic membrane proteins of phosphate binding protein-dependent permeases

In a previous report, Saurin et al. compared 61 integral membrane proteins from 35 uptake systems and defined ligand specific signatures identifying functionally related inner membrane proteins of binding protein-dependent bacterial permeases [35]. These are located within a hydrophilic segment between the penultimate and ante-penultimate potential transmembrane segment in a previously described conserved region found in integral membrane proteins of all bacterial binding protein-dependent transport systems [29,36,37]. For the phosphate-specific transport systems, only two components were known at that time (*E. coli* PstA and PstC) and a consensus sequence could not be defined. After the identification of the PstC-1, PstC-2, PstA-1 and PstA-2 components in *M. tuberculosis* and extensive homology searches in databases, 17 bacterial inner membrane components of phosphate permeases were identified. The proteins originated from Gram-negative bacteria (*Escherichia coli*, *Pseudomonas aeroginosa*, *Haemophilus influenzae*), Gram-positive bacteria (*Bacillus subtilis*), cyanobacteria (*Synechocystis* spp.), mycoplasma (*Mycoplasma genitalium*) and mycobacteria (*Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium intracellulare*). Peptide segments fitting the conserved region [29] were aligned (Table 1) and a highly conserved degenerated motif of 25 amino acids was defined: [P,S]-X(3)-[R,K,I]-X-[A,G,S]-[A,S,C]-X-[A,G]-[L,I,V]-G-X-[T,S,P,Y]-X(4)-[I,V,L,-A]-X(2)-[I,V,T]-X(2)-[P,K,R]. When used to search protein sequences in a comprehensive non-redundant database (PROT), this motif selected specifically all the sequences presented in Table 1. Moreover, a hydrophobic protein belonging to the phosphate transport system of *Mycoplasma capricolum*, but not included in our set of test sequences, was also specifically found (accession number S46902). None of the hydrophobic membrane proteins of other transport systems were detected. The motif defined here may therefore be considered as a signature for the inner membrane components of bacterial phosphate transporters.

4. Discussion

Two new *M. tuberculosis* genes, *pstS-3* and *pstC-2*, coding for putative phosphate permease subunits have been cloned and sequenced. The deduced molecular weights of these gene products are 37.731 kDa and 34.136 kDa respectively. PstS-3

is a putative phosphate binding protein highly similar to *E. coli* PstS. *M. tuberculosis* PstC-2 corresponds to a hydrophobic protein with six putative transmembrane helices, homologous to the transmembrane *E. coli* PstC protein [7]. Together with the previously identified homolog of the *E. coli pstA*, now called *pstA-1* [15], these two *M. tuberculosis* genes constitute a new cluster of three genes distant in about 3 kb from another putative phosphate permease operon we described before [14].

Using one-step affinity chromatography purification of *M. bovis* BCG culture filtrate, we obtained the previously described Ag 88 [15] with a degree of purity allowing protein sequencing and thus demonstrated that this protein is encoded by the *pstS-3* gene.

We defined a pattern identifying inner membrane proteins of phosphate binding protein-dependent permeases. It is located in a previously described conserved region found in integral membrane proteins of all bacterial binding protein-dependent transport systems [29,36,37] and may be used for the identification of substrate specificity of new tranport systems. As suggested by Saurin et al., the diversity found within the conserved region of transmembrane proteins of bacterial tranporters may be required for a ligand-specific recognition of highly conserved ATP-binding proteins [35], i.e. PstB in the case of phosphate transport systems.

Previous findings [10,14,15] together with the results reported here suggest that two or three complex phosphate permease systems may exist in *M. tuberculosis*. The identification of genes encoding several inner membrane permease components, i.e. two PstC and two PstA subunits, but only one ATP-binding PstB subunit so far, has led us to the following hypothesis. Since at least three phosphate binding proteins equivalent to the *E. coli* PstS exist in mycobacteria, it is possible that at least two functional transmembrane complexes, each composed of a specific PstA and PstC, are specifically adapted either to PstS-1 or to PstS-2 and PstS-3 (PstS-2 and PstS-3 displaying 94% similarity between themselves but only 78% and 73% similarity with PstS-1 respectively). The same PstB subunit could energize all the different systems by interacting with different PstA-PstC complexes. The possible composition of the multi-subunits complexes could be PstS-1.PstA-2.PstC-1.PstB, PstS-2.PstA-1.PstC-2.PstB and PstS-3.PstA-1.PstC-2.PstB. It is conceivable that certain regions of the cytoplasmic surface of the two integral membrane components (such as the signature containing loop) are more conserved than the external loops presumably recognized by the ligand binding component. These two suggestions are further supported by the observation that the genes for the various ATP-binding subunits are highly conserved even among permeases transporting different ligands [38].

In *M. leprae*, all these phosphate transport genes were found in two different non-adjacent cosmids: B1308 for *pstB*, *pstS-1*, *pstC-1*, *pstA-2* and *pstS-2* and B650 [39] for *pstS-3*, *pstC-2* and *pstA-1* (Fig. 1). Moreover, one additional *M. leprae* cosmid, cosmid B2266 [39], contains genes coding for phoS, phoW, phoX, phoT and phoY proteins homologous to *E. coli* PstS, PstC, PstA, PstB and PhoU proteins, respectively. Therefore it is conceivable that genes coding for a fourth phosphate transport permease are also present in the *M. tuberculosis* genome. The probable existence of three or four phosphate transporters in mycobacteria seems to be new among known prokaryotes. So far, genes coding for only one phosphate permease have been identified in several bacterial genomes, including the completely sequenced genomes from *Haemophilus influenzae* [25] and *Mycoplasma genitalium* [40]. The reason for such duplications in mycobacteria is unknown but a possible explanation could be that the different permeases are adapted to various related ligands (organic or inorganic phosphate, phosphonate) or to different external environments (pH, phosphate concentrations). In this respect, it could be of interest to study the expression of the genes coding for the different permeases in infected macrophages or in different growth conditions.

Progress in studying the structure and function of the phosphate binding subunits (PstS-1, PstS-2 and PstS-3) could lead to the screening and design of new antimycobacterial drugs.

Acknowledgements: We thank Dr. R.A. Young for the *M. tuberculosis* λgt11 library. Competent bio-computing support was provided by the members of the Belgian EMBnet Node. This work was supported by grants 3.4543.95 from FRSM (Fund for Medical Scientific Research, Belgium). We thank the 'Loterie Nationale' for the acquisition of an automatic sequencer.

References

[1] Bloom, B.R.C. and Murray, C.J.L. (1992) Science 257, 1055–1064.
[2] Kaufmann, S.H.E. and Van Embden, J.D.A. (1993) Trends Microbiol. 1, 2–5.
[3] Ames, G.F.-L. (1993) Int. Rev. Cytol. 137, 1–35.
[4] De la Salle, H. (1994) Science 265, 237–241.
[5] Perego, M., Higgins, C.F., Pearce, S.R., Gallagher, M.P. and Hoch, J.A. (1991) Mol. Microbiol. 5, 173–185.
[6] Gilson, E., Alloing, G., Schmidt, T., Claverys, J.P., Dudler, R. and Hofnung, M. (1988) EMBO J. 7, 3971–3974.
[7] Surin, B.P., Rosenberg, H. and Cox, G.B. (1985) J. Bacteriol. 161, 189–198.
[8] Torriani, A. (1990) Bioessays 12, 371–376.
[9] Wanner, B.L. (1993) J. Cell. Biochem. 51, 47–54.
[10] Andersen, A.B. and Hansen, E.B. (1989) Infect. Immun. 57, 2481–2488.
[11] Harboe, M. and Wiker, H.G. (1992) J. Infect. Dis. 166, 874–884.
[12] Young, D.B. and Garbe, T.R. (1991) Res. Microbiol. 142, 55–65.
[13] Chang, Z., Choudhary, A., Lathigra, R. and Quiocho, F.A. (1994) J. Biol. Chem. 269, 1956–1958.
[14] Braibant, M., Lefèvre, P., De Wit, L., Peirs, P., Ooms, J., Huygen, K., Andersen, A.B. and Content, J. (1996) Gene, in press.
[15] Braibant, M., De Wit, L., Peirs, P., Kalai, M., Ooms, J., Huygen, K. and Content, J. (1994) Infect. Immunol. 62, 849–854.
[16] Huygen, K., Drowart, A., Harboe, M., Ten-Berg, R., Cogniaux, J. and Van Vooren, J.P. (1993) Infect. Immunol. 61, 2687–2693.
[17] Huygen, K., Ljungqvist, L., Ten-Berg, R. and Van Vooren, J.P. (1990) Infect. Immunol. 58, 2192–2197.
[18] Young, R.A., Bloom, B.R., Grosskinsky, C.M., Ivanyi, J., Thomas, D. and Davis, R.W. (1985) Proc. Natl. Acad. Sci. USA 82, 2583–2587.
[19] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) in Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.
[20] Laemmli, U.K. (1970) Nature 227, 680–685.
[21] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[22] Stone, L.K. and Williams, K.R. (1993) in A Practical Guide to Protein and Peptide Purification of Microsequencing (Matsudaira, P., Ed.) pp. 43–54. Academic Press, CA.
[23] Marck, C. (1988) Nucleic Acids Res. 16, 1829–1836.
[24] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[25] Fleischmann, R.D. et al. (1995) Science 269, 496–512.
[26] Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. and Tabata, S. (1995) DNA Res. 2, 153–166.
[27] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.

[28] Sander, C. and Schneider, R. (1991) Proteins 9, 56–68.
[29] Bairoch, A. (1991) Nucleic Acids Res. 25, 2241–2245.
[30] Gafvelin, G. and von Heijne, G. (1994) Cell 77, 401–412.
[31] Webb, D.C., Rosenberg, H. and Cox, G.B. (1992) J. Biol. Chem. 267, 24661–24668.
[32] Hopkins, C.M., White, F.F., Heaton, L.A., Guikema, J.A. and Leach, J.E. (1995) DNA Sequence 5, 299–305.
[33] Scanlan, D.J., Mann, N.H. and Carr, N.G. (1993) Mol. Microbiol. 10, 181–191.
[34] Luecke, K. and Quiocho, F.A. (1990) Nature 347, 402–406.
[35] Saurin, W., Köster, W. and Dassa, E. (1994) Mol. Microbiol. 12, 993–1004.
[36] Dassa, E. and Hofnung, M. (1985) EMBO J. 4, 2287–2293.
[37] Kerppola, R.E. and Ames, G.F. (1992) J. Biol. Chem. 267, 2329–2336.
[38] Higgins, C.F., Gallagher, M.P., Mimmak, M.L. and Pearce, S.R. (1988) BioEssays 8, 111–116.
[39] Eiglmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T. (1993) Mol. Microbiol. 7, 197–206.
[40] Fraser, C.M. et al. (1995) Science 270, 397–403.
[41] Cox, G.B., Webb, D. and Rosenberg, H. (1989) J. Bacteriol. 171, 1531–1534.
[42] Kato, J., Nikata, T. and Ohtake, H. (1994) J. Bacteriol. 176, 5874–5877.