Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 790 (2003) 337–348
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

High-level expression of the mycobacterial porin MspA in *Escherichia coli* and purification of the recombinant protein

Christian Heinz, Stefanie Karosi[1], Michael Niederweis*

Lehrstuhl für Mikrobiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Staudtstrasse 5, D-91058 Erlangen, Germany

Abstract

MspA is the prototype of a new family of tetrameric porins and provides the main general diffusion pathway for hydrophilic compounds through the outer membrane of *Mycobacterium smegmatis*. Structural analysis was hampered by the scarce amount of pure protein. After replacement of the GC-rich codons of the *mspA* gene by codons optimal for high-level expression in *Escherichia coli*, the mature MspA protein was overproduced in *E. coli*. The recombinant MspA (rMspA) monomer (M_r 20 000) was purified by anion exchange and hydrophobic interaction chromatography yielding 2.6 mg pure protein per liter of culture. This exceeded the yield of the native protein 10-fold. Circular dichroism revealed that rMspA is folded in a native-like structure. rMspA assembled partially to the channel-forming tetramer both during expression in *E. coli* and after purification in vitro. Thus, overexpression in *E. coli* and chromatographic purification are key steps towards a high resolution structure of MspA.

Keywords: *Escherichia coli*; *Mycobacterium smegmatis*; Genes; Proteins

1. Introduction

The mycobacterial cell envelope is unique among bacteria. It consists of a cytoplasma membrane and an outer membrane, whose inner leaflet is built from very long chain fatty acids, the mycolic acids, that are covalently attached to an arabinogalactan–peptidoglycan co-polymer. The mycolic acid membrane is thought to be about twice as thick as the outer membrane of Gram-negative bacteria and resembles its function by forming an efficient permeability barrier [1]. The diffusion of small and hydrophilic solutes across this outer membrane is mediated by channel-forming proteins, the porins [2]. MspA was identified as an oligomeric porin of *Mycobacterium smegmatis* composed of subunits with a relative molecular mass of 20 000 [3]. Enzyme-linked immunosorbent assays and immunogold labeling experiments demonstrated that MspA is localized in the cell wall of *M. smegmatis* [4]. Uptake of glucose by an *mspA* deletion mutant was fourfold slower compared to the wild-type indicating that MspA is the main porin of *M. smegmatis* [4]. MspA is the prototype of a family of four very similar porins of *M. smegmatis*, which did not show any homology to any other known protein [4]. Electron microscopy of MspA in the cell walls of *M. smegmatis* in combination with crosslinking experiments revealed a tetrameric pore with a central channel of 10 nm length

*Corresponding author. Tel.: +49-9131-852-8989; fax: +49-9131-852-8082.
E-mail address: mnieder@biologie.uni-erlangen.de (M. Niederweis).
[1]Present address: Nikolaus-Fiebiger Center for Molecular Medicine, Glückstrasse 6, D-91054 Erlangen, Germany.

1570-0232/03/$ – see front matter
doi:10.1016/S1570-0232(03)00130-2

and an inner diameter (I.D.) of 2.5 nm [5]. These structural features are unique and clearly distinguish the MspA pore and related mycobacterial porins from all other known bacterial porins [6]. Infrared and circular dichroism spectroscopy revealed that the MspA tetramer consists mainly of antiparallel β-strands most likely organized in a channel-forming β-barrel of remarkable stability. Heating to 92 and 112 °C was required to dissociate the MspA tetramer and to unfold the β-sheet domain in the monomer, respectively [7]. The stability of the MspA tetramer exceeded the stability of the porins of Gram-negative bacteria for every condition tested and was not reduced in the presence of 2% sodium dodecyl sulfate (SDS) and at any pH from 0 to 14 [7].

However, milligram quantities of protein are needed for further structural studies by NMR spectroscopy or X-ray crystallography. This can hardly be achieved by purification of MspA from *M. smegmatis*, which yielded about 230 μg pure protein per liter of culture using a recently improved protocol [8]. Since overexpression systems are not available for mycobacteria [9], heterologous production of mycobacterial proteins seems to be the only alternative. However, expression of mycobacterial genes in *Escherichia coli* is often hampered by the different codon usages of both organisms as reflected by their G+C contents of about 51% [10] and about 68% [11], respectively. This limitation can be overcome by converting codons of mycobacterial genes with a low usage in *E. coli* to high-usage codons [12].

In this study, we describe the overexpression of *mspA* from *M. smegmatis* in *E. coli* after synthesis of the gene with codons preferentially used in highly expressed *E. coli* genes. The MspA protein was purified in milligram amounts using three chromatographic steps. This is the first study reporting the heterologous overexpression of a mycobacterial cell wall protein and facilitates the structural analysis of MspA as the prototype of a new family of outer membrane channel proteins.

2. Experimental

2.1. Materials

Chemicals were obtained from Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany) or Sigma (Munich, Germany) at the highest purity available. Enzymes for DNA restriction and modification were from New England Biolabs (Schwalbach, Germany), Boehringer (Mannheim, Germany), Stratagene (Heidelberg, Germany) and Pharmacia (Freiburg, Germany). Oligonucleotides were obtained from PE Applied Biosystems (Weiterstadt, Germany).

2.2. Bacterial strains and growth conditions

Escherichia coli DH5α was used for all cloning experiments, and *E. coli* BL21(DE3) for overexpression of recombinant (r) MspA. These strains were routinely grown in LB medium [13] at 37 °C. Kanamycin was used when required at concentrations of 30 μg/ml.

2.3. Synthesis of the synmspA gene

Assembly polymerase chain reaction (PCR) is a method for the synthesis of long DNA sequences from large numbers of oligonucleotides and was carried out as described [14]. Both strands of the *synmspA* gene were covered by oligonucleotides of 40 nucleotides (nt) length, which overlapped over 20 nt (Table 1). The oligonucleotides were mixed to a final concentration of 50 fmol/μl for each oligonucleotide. The assembling reaction was carried out in a volume of 20 μl containing Taq buffer (50 m*M* KCl, 1.5 m*M* $MgCl_2$, 10 m*M* Tris–HCl, pH 9 at room temperature), deoxyribose nucleotide triphosphates (dNTPs; 0.25 m*M* for each nucleotide) and a mixture of Taq and Pfu polymerase (1 and 0.025 U/μl, respectively) using the following temperature profile: 2 min at 95 °C, 55× (30 s at 95 °C, 30 s at 50 °C, 2 min at 72 °C), 5 min at 72 °C. This sample was diluted 500-fold and used as a template for amplification of the *synmspA* gene with the oligonucleotides MSPAF01 and MSPAR01 (Table 1).

2.4. Construction of plasmids

The *mspA* gene encoding the mature MspA without leader sequence was amplified from the plasmid pPOR6 [3] with the oligonucleotides mspA-fwd (5′–ATATA*GAATTC*AAG AAGGAG ATATACAT **ATG**-GGCCTGGACAACGAGCTGAGCCTC–3′), which introduced an *Eco*RI restriction site (italic), a

Table 1
Sequences of the oligonucleotides used for PCR assembly of the *synmspA* gene

Oligonucleotide	Sequence (5′–3′)
MSPAF01	GTATACATATGGGCCTGGAC
MSPAF02	AACGAACTGTCCCTGGTTGACGGCCAGGACCGTACCCTGA
MSPAF03	CCGTTCAGCAGTGGGACACCTTCCTGAACGGTGTTTTCCC
MSPAF04	GCTGGACCGTAACCGTCTGACCCGTGAATGGTTCCACTCC
MSPAF05	GGTCGTGCGAAATACATCGTTGCGGGTCCGGGTGCGGACG
MSPAF06	AGTTCGAAGGTACCCTGGAACTGGGTTACCAGATCGGCTT
MSPAF07	CCCGTGGTCCCTGGGTGTTGGTATCAACTTCTCTTACACC
MSPAF08	ACCCCGAACATCCTGATCGACGACGGTGACATCACCGCTC
MSPAF09	CGCCGTTCGGTCTGAACTCTGTTATCACCCCGAACCTGTT
MSPAF10	CCCGGGTGTTTCTATCTCTGCTGATCTGGGCAACGGTCCG
MSPAF11	GGTATCCAGGAAGTTGCTACCTTCTCTGTAGACGTCTCTG
MSPAF12	GTGCTGAAGGTGGTGTTGCTGTTTCTAACGCTCACGGCAC
MSPAF13	CGTTACCGGTGCGGCTGGCGGTGTTCTGCTGCGTCCGTTC
MSPAF14	GCTCGTCTGATCGCTTCTACCGGTGACTCTGTTACCACCT
MSPAF15	ACGGTGAACCGTGGAACATGAACTGAAAGCTTGAGAG
MSPAR01	CTCTCAAGCTTTCAGTTCATG
MSPAR02	TTCCACGGTTCACCGTAGGTGGTAACAGAGTCACCGGTAG
MSPAR03	AAGCGATCAGACGAGCGAACGGACGCAGCAGAACACCGCC
MSPAR04	AGCCGCACCGGTAACGGTGCCGTGAGCGTTAGAAACAGCA
MSPAR05	ACACCACCTTCAGCACCAGAGACGTCTACAGAGAAGGTAG
MSPAR06	CAACTTCCTGGATACCCGGACCGTTGCCCAGATCAGCAGA
MSPAR07	GATAGAAACACCCGGGAACAGGTTCGGGGTGATAACAGAG
MSPAR08	TTCAGACCGAACGGCGGAGCGGTGATGTCACCGTCGTCGA
MSPAR09	TCAGGATGTTCGGGGTGGTGTAAGAGAAGTTGATACCAAC
MSPAR10	ACCCAGGGACCACGGGAAGCCGATCTGGTAACCCAGTTCC
MSPAR11	AGGGTACCTTCGAACTCGTCCGCACCCGGACCCGCAACGA
MSPAR12	TGTATTTCGCACGACCGGAGTGGAACCATTCACGGGTCAG
MSPAR13	ACGGTTACGGTCCAGCGGGAAAACACCGTTCAGGAAGGTG
MSPAR14	TCCCACTGCTGAACGGTCAGGGTACGGTCCTGGCCGTCAA
MSPAR15	CCAGGGACAGTTCGTTGTCCAGGCCCATATGTATAC

ribosome binding site (underlined) and a start codon (bold), and mspA-rev (5′–TATAT*AAGCTT*CGAA-CGGCGGTCCAGGAATCAG–3′), which intro-duced a *Hin*dIII restriction site (italic). The sequence between the *Eco*RI site and the start codon is identical to that in all translational fusion vectors of the pET series (Novagen, Germany). The PCR fragment containing the *mspA* gene was digested with *Hin*dIII and *Eco*RI and cloned into the plasmid pET24+ (Novagen, Germany) to give pMN500. The PCR fragment containing the *synmspA* gene was digested with *Nde*I and *Hin*dIII and cloned into pMN500 resulting in the plasmid pMN501. The correct sequences of both genes were verified by cycle sequencing using an ABI PRISM 310 Genetic Analyzer and a protocol provided by the manufac-turer (PE Applied Biosystems).

2.5. Expression analysis in E. coli

E. coli BL21(DE3) transformed with pMN500 or pMN501 was grown in selective LB medium to an A_{600} of 0.5 and induced with a final concentration of 1 m*M* isopropyl β-D-thiogalactoside (IPTG). At different times samples were withdrawn from the culture. The cells were lysed by boiling and cen-trifuged to remove the cell debris. Samples were diluted fourfold with loading buffer [140 m*M* Tris–HCl, pH 7.5, 30% (w/v) glycerol, 4% (w/v) SDS, 0.1% (w/v) bromophenol blue] and incubated for 10 min at room temperature before gel electrophoresis.

2.6. Solubility of recombinant MspA (rMspA)

E. coli cells BL21(DE3) containing the plasmid

pMN501 or the parent vector pET24+ were induced with IPTG as described (see Section 2.5) and harvested after 4 h. The cells were resuspended in CLP buffer (2 *M* NaCl, 0.2 *M* Tris–HCl, pH 8.0) to give a final A_{600} of 50 and sonicated two times for 10 s (cycle: 0.9, 45 W) and centrifuged at 16 000 *g* and 4 °C. The supernatants were then centrifuged at 100 000 *g* for 1 h at 4 °C. The supernatants and the pellets were mixed with loading buffer (140 m*M* Tris–HCl pH 7.5, 30% (w/v) glycerol, 4% (w/v) SDS, 0.1% (w/v) bromophenol blue), boiled for 2 min and equivalent amounts of protein were loaded on a denaturing 10% polyacrylamide gel.

2.7. Purification of recombinant MspA (rMspA)

Two liters of an *E. coli* BL21(DE3) culture were grown to an A_{600} of 0.5, induced with a final concentration of 1 m*M* IPTG and harvested 5 h after induction. 6.8 g cells (wet mass) were suspended in 20 ml buffer A [25 m*M* 4-(2-hydroxyethyl)-1-piperazineethane sulfonic acid (HEPES), pH 7.5]. The suspension was sonicated four times for 40 s (cycle: 0.9, power: 70%) using the sonicator Sonoplus UW 2070 (Bandelin Electronic, Berlin, Germany) with the sonotrode VS70 to lyse the cells. The suspension was centrifuged at 27 000 *g* and 4 °C for 15 min to remove cell debris. The supernatant was further purified from remaining insoluble particles by centrifugation at 55 000 rpm in a Ti60 rotor (Beckman) and 4 °C for 2 h and then loaded on an anion-exchange column Poros HQ/M (100×16 mm, Perseptive Biosystems, Cambridge, MA, USA). The column was washed with 100 ml buffer A. Bound proteins were eluted with a gradient from 100% buffer A to 100% buffer B (25 m*M* HEPES, pH 7.5, 2 *M* NaCl) over 200 ml. Sixty-two fractions of 6 ml were collected. The fractions were analyzed using denaturing polyacrylamide gels which were stained with Coomassie Blue. Two fractions (31+32) with the highest rMspA content were pooled and adjusted to an ammonium sulfate concentration of 10% of a saturated solution. This mixture was incubated for 30 min on ice. After centrifugation at 27 000 *g* and 4 °C for 15 min, the supernatant was loaded on a hydrophobic interaction chromatography (HIC) column Poros 20PE (100×16 mm, Perseptive Biosystems). The column was washed with 40.2 ml buffer AAS (25 m*M* HEPES, pH 7.5, 10% ammonium sulfate of a saturated solution). Bound proteins were eluted with a gradient from 100% buffer AAS to 100% buffer A over 60.3 ml. Seventy-three fractions of 3 ml were collected. The fractions were analyzed using denaturing polyacrylamide gels which were stained with Coomassie Blue. Six fractions (39–44) with the highest rMspA content were pooled and adjusted to an ammonium sulfate concentration of 40% of a saturated solution. This mixture was incubated for 30 min on ice. After centrifugation at 27 000 *g* and 4 °C for 15 min, the precipitated protein was dissolved in 800 μl buffer A and centrifuged at 4 °C for 5 min to remove insoluble material. The protein solution was dialyzed against 500 ml buffer A at 4 °C for 12 h (molecular mass cut off: 2000). The retentate was loaded on a size exclusion column G4000PW_{XL} (300×7.8 mm, TosoHaas, Stuttgart, Germany). Proteins were eluted with 28.7 ml buffer A at a flow-rate of 0.5 ml/min. Twenty-nine fractions were collected and analyzed using denaturing polyacrylamide gels, which were stained with Coomassie Blue. Two fractions (12+13) containing apparently pure rMspA were pooled. The volume of the pooled fractions was reduced from 2 to 0.2 ml by ultrafiltration using a filtron device (molecular mass cut off: 30 000, membrane type: Omega, Filtron Technology, Northborough, MA, USA). The protein concentration was determined as described below.

2.8. Polyacrylamide gel electrophoresis (PAGE)

Protein samples were analyzed in denaturing polyacrylamide gels as described earlier [8].

2.9. Determination of protein concentrations

Protein concentrations were determined using bicin choninic acid (BCA) [15] and bovine serum albumin (BSA) as standard protein, according to the manufacturer's instruction (BCA protein assay kit, Pierce, Rockford, IL, USA).

2.10. Immunoblot analysis

Increasing amounts of MspA and rMspA were loaded on denaturing 10% polyacrylamide gels and

blotted onto a nitrocellulose membrane. Proteins were visualized using the MspA antiserum [3] and a chemiluminescence reaction (ECLplus kit, Amersham Pharmacia, Vienna, Austria).

2.11. Circular dichroism spectroscopy

Circular dichroism (CD) spectra were measured on a Jasco J-715 spectropolarimeter (Japan Spectroscopic Co.) using a protein concentration of 0.26 mg/ml in NaP buffer (25 m*M* NaH_2PO_4/Na_2HPO_4, pH 7.5) and an optical path length of 1 mm. Buffer baselines were measured under identical conditions and subtracted from the corresponding spectrum. The raw data were transformed to mean residue molar ellipticity and smoothed using the standard analysis program (Jasco).

3. Results

3.1. Overexpression of the mspA gene in E. coli

The pET system relies on the selective transcription of the gene of interest by the bacteriophage T7 RNA polymerase [16] and is one of the most widely used systems for expression of recombinant proteins in *E. coli* [17]. To produce the porin MspA from *M. smegmatis* in *E. coli*, the *mspA* gene encoding the mature MspA protein without signal sequence was cloned as a transcriptional fusion with the strong late promoter of the bacteriophage T7 into the plasmid pET24+ to give the plasmid pMN500. However, induction of *E. coli* cells containing pMN500 with IPTG did not lead to detectable expression of the *mspA* gene, although the *E. coli lacZ* gene was overexpressed using the same expression signals in the control vector pET28+lacZ under the same experimental conditions (data not shown). We assumed that inefficient translation of the *mspA* mRNA might have caused the expression problems, since the *mspA* gene contained nine and 16 codons, which are used with a frequency of less than 10 and 20%, respectively, in highly expressed *E. coli* genes [18]. This included four CCC codons for proline, which are known to cause translational problems in *E. coli* [19]. Therefore, the *mspA* gene was synthesized in vitro by assembly PCR [14] from 30 oligonucleotides. All codons of the *synmspA* gene are preferentially used in highly expressed *E. coli* genes [18] requiring 71 sequence changes compared to the native *mspA* gene (Fig. 1). This gene was cloned into pMN500 to replace the *mspA* gene resulting in the plasmid pMN501 (Fig. 2). Addition of IPTG to *E. coli* BL21(DE3) cells containing pMN501 induced the expression of two additional proteins with apparent relative molecular masses of 20 000 and 100 000 compared to uninduced cells and cells containing the parent plasmid pET24+, which did not contain an *mspA* gene (Fig. 3A). In a previous work, these proteins were purified by preparative gel electrophoresis and identified as the monomeric and oligomeric form of MspA [3]. The tetrameric form showed the same channel activity in lipid bilayer experiments as MspA purified from *M. smegmatis* [3]. Quantitative image analysis using the software NIH image 1.62 (http://rsb.info.nih.gov/nih-image/) revealed that about 15% of rMspA was present in the tetrameric form (Fig. 3A, lane 4). The MspA monomer was strongly expressed at 60 min up to at least 5 h after induction in a time-dependent expression analysis by gel electrophoresis (data not shown). To analyze whether rMspA was soluble, the supernatant of *E. coli* cells lysed after induction of *mspA* expression was centrifuged for 1 h at 100 000 *g*. The major fraction of rMspA was found in the supernatant indicating that rMspA was mainly expressed in a water-soluble form (Fig. 3B).

3.2. Purification of recombinant MspA (rMspA)

For large-scale purification of proteins, chromatography offers many advantages such as different separation principles, high speed, little hands-on time and the option to automate purification. Therefore, we decided to purify rMspA chromatographically instead of the previously used preparative gel electrophoresis [3]. Expression of *synmspA* was induced with IPTG in a 2-l culture of *E. coli* BL21(DE3) containing the plasmid pMN501 (6.8 g cells wet mass) were harvested 5 h after induction. Since the amount of monomeric rMspA exceeded that of tetrameric rMspA about sixfold (see above; Fig. 3A, lane 4), monomeric rMspA was purified following the procedure consisting of three chromatographic steps. The cell were lysed by ultrasonication and the

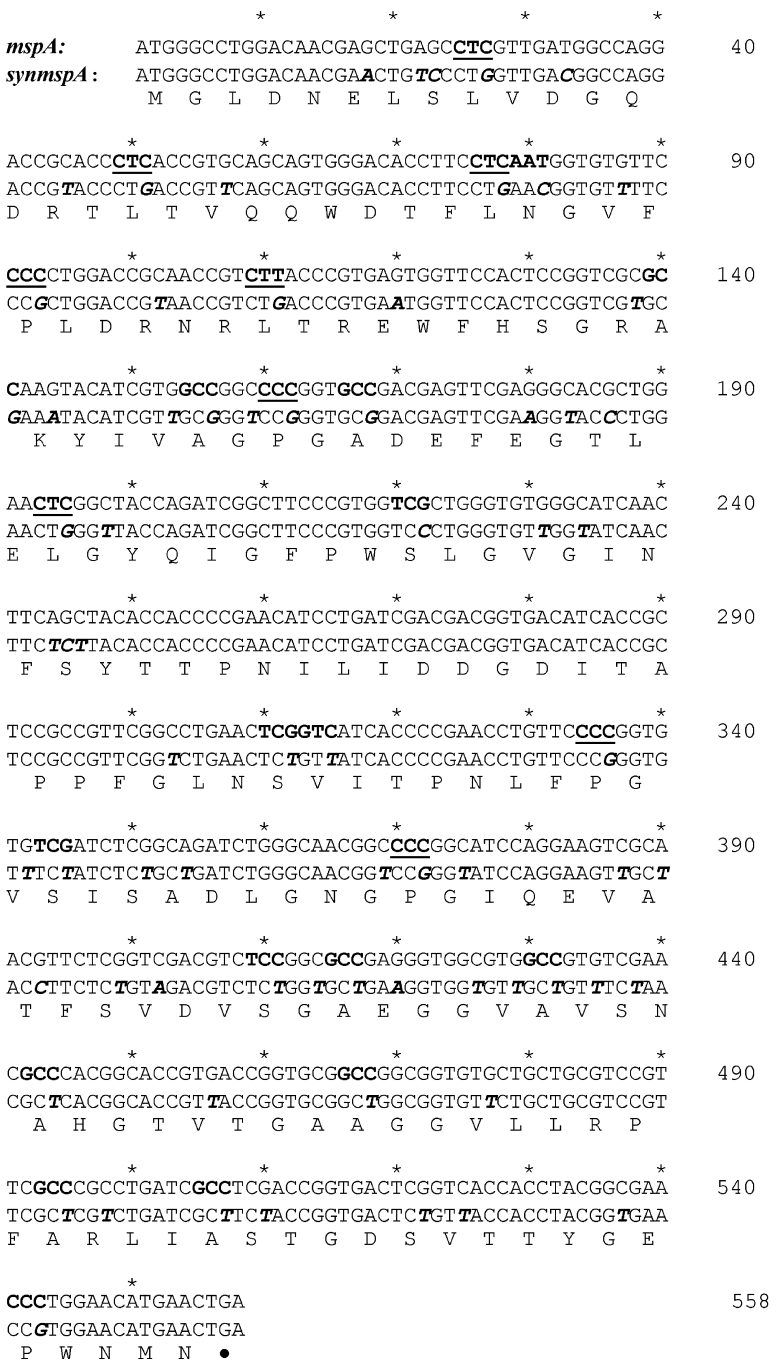

Fig. 1. Adaptation of the codon-usage of the *mspA* gene to *E. coli*. The upper sequence shows the *mspA* gene from *M. smegmatis* as cloned into the expression vector pMN500. The lower sequence represents the *synmspA* sequence, which contains codons preferentially used by highly expressed *E. coli* genes [18]. Both genes encode for the mature MspA protein. The signal sequences were replaced by methionine. The amino acid sequence of MspA is depicted below the corresponding codons in a one letter code. The frequency of boldfaced and boldfaced/underlined codons of the native *mspA* gene is less than 20 and 10%, respectively, in strongly expressed *E. coli* genes [18]. The sequence alterations in the *synmspA* gene compared to the original *mspA* gene are marked by italic and boldfaced letters.

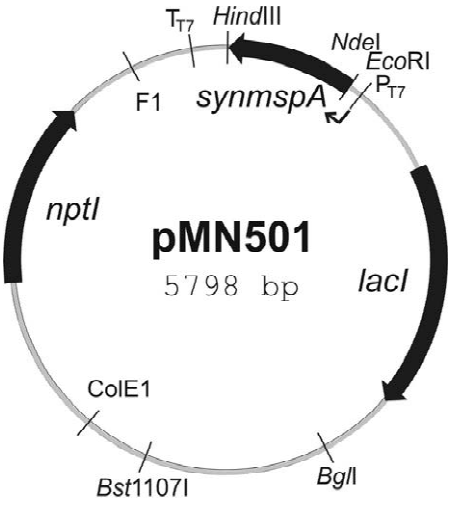

Fig. 2. Restriction map of the *mspA* overexpression vector. The *synmspA* gene of pMN501 is transcribed from the P_{T7} promoter under control of the *lac* operator. The *lacI* gene encodes for the lactose repressor. The *nptI* gene confers resistance to kanamycin. The symbol T_{T7} indicates a transcription terminator of the bacteriophage T7. The depicted restriction sites are singular. The plasmid contains the origins *ColE1* and *F1* for replication in *E. coli* and of phage f1.

lysate containing 225 mg protein was loaded on an anion-exchange column. Recombinant MspA eluted in a sharp peak at 510 m*M* NaCl during gradient elution (Fig. 4A). The two fractions indicated by a bracket contained mainly rMspA and were pooled for further purification. The protein content was 39.3 mg.

Different amounts of ammonium sulfate were added to rMspA to determine the ammonium sulfate concentration suitable for hydrophobic interaction chromatography (HIC). Recombinant MspA precipitated at an ammonium sulfate concentration of 15% of a saturated solution or above (data not shown). Therefore, the pooled rMspA fractions were adjusted to an ammonium sulfate concentration of 10% of a saturated solution. Most of the contaminating proteins did not bind to the phenyl ether ligands of the HIC column and were eluted in the breakthrough (Fig. 4B, first peak). In contrast, rMspA bound to the HIC column under those conditions and eluted at an ammonium sulfate concentration of 3.8% of a saturated solution after applying an inverse gradient (Fig. 4B, second peak). The six fractions indicated by a bracket were pooled for further purification (Fig.

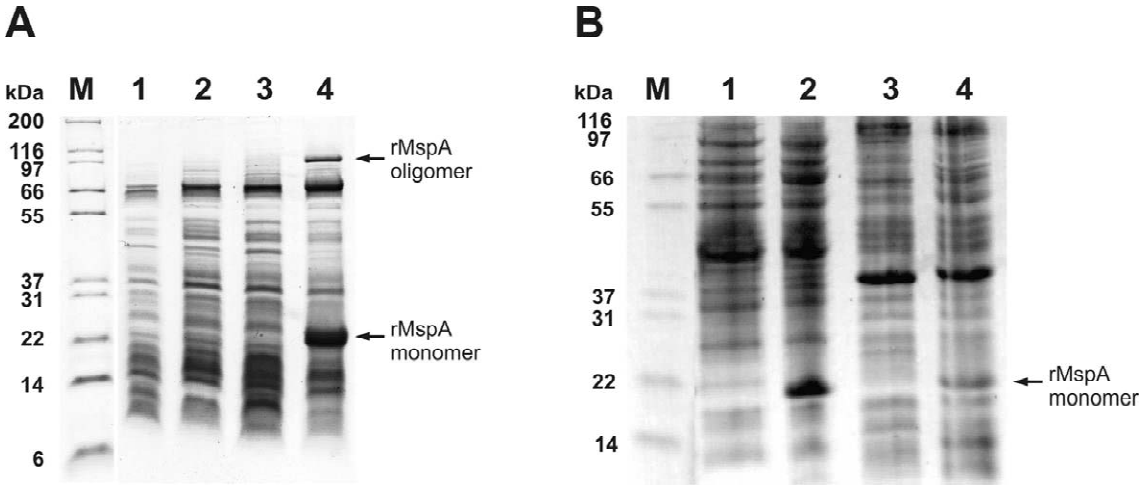

Fig. 3. Overexpression of *synmspA* in *E. coli*. (A) The supernatants of cells of *E. coli* BL21(DE3) containing the parent vector pET24+ or the *mspA* expression vector pMN501 before and after induction with IPTG were analyzed in a 10% denaturing polyacrylamide gel stained with Coomassie. Lanes: M, molecular mass marker; 1, pET24+ before induction; 2, pET24+ 5 h after induction; 3, pMN501 before induction; 4, pMN501 5 h after induction. (B) Solubility of overexpressed rMspA. *E. coli* BL21(DE3) cells containing pMN501 and the parent vector pET24+ were induced with IPTG and lysed 4 h after induction. The lysate was analyzed by ultracentrifugation followed by gel electrophoresis in a 10% denaturing polyacrylamide gel stained with Coomassie. In all lanes, similar amounts of protein were loaded. Lanes: M, molecular mass marker; 1, supernatant after ultracentrifugation, pET24+; 2, supernatant after ultracentrifugation, pMN501; 3, pellet after ultracentrifugation, pET24; 4, pellet after ultracentrifugation, pMN501.

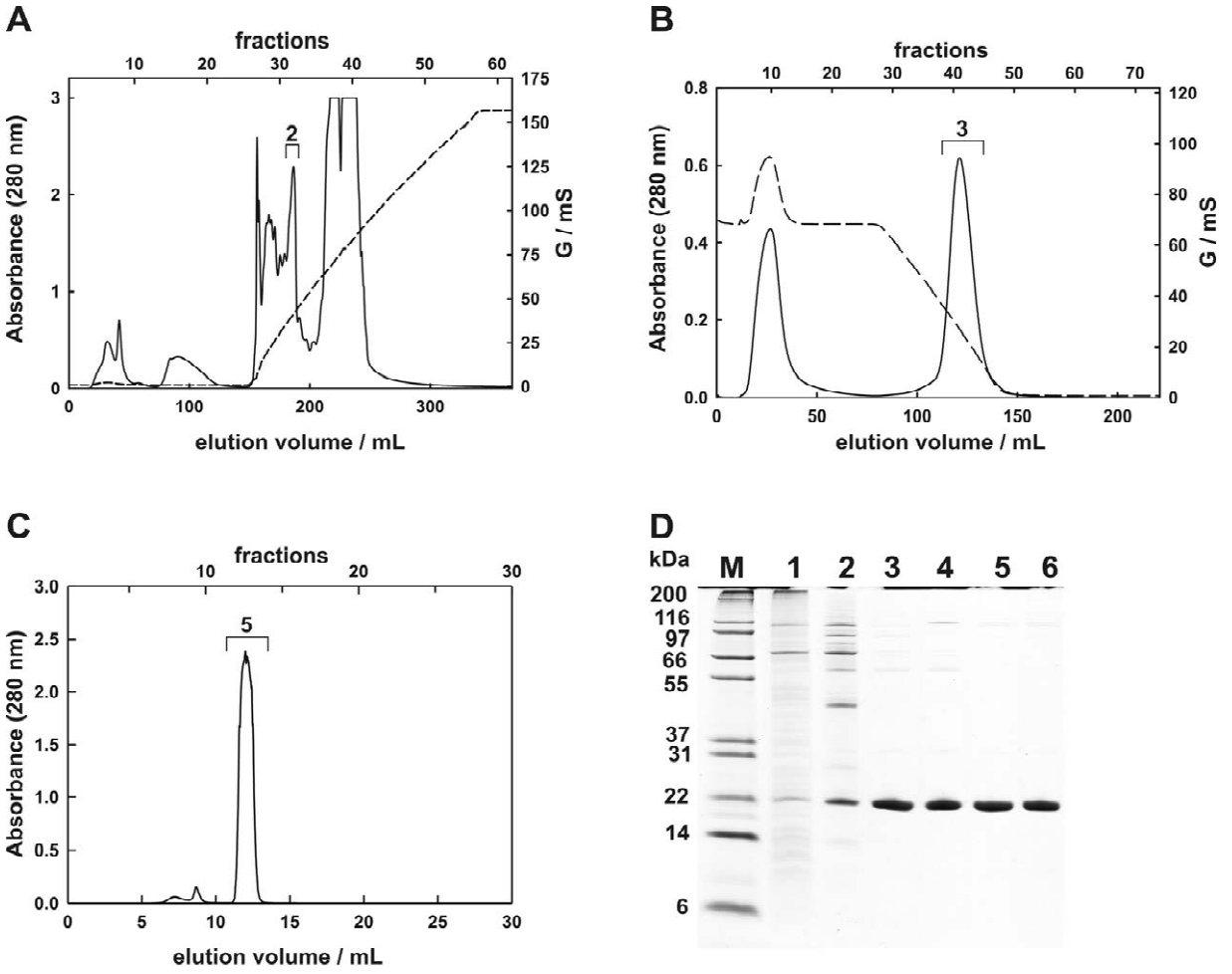

Fig. 4. Large-scale purification of recombinant MspA. rMspA was purified in three consecutive chromatographic steps. In all chromatograms, the solid and the dashed lines represent the absorbance at 280 nm and the conductivity, respectively. The fractions, which were pooled for further purification are marked with a bracket. The numbers above the brackets indicate the lanes in the gel of the final electrophoretic analysis of all purification steps (D). (A) Anion-exchange chromatography (AEC). rMspA eluted at 0.51 *M* NaCl. Fractions 31+32 were pooled. (B) Hydrophobic interaction chromatography (HIC). rMspA eluted at an ammonium sulfate concentration of 3.8% of a saturated solution. Fractions 39 to 44 were pooled. (C) Size exclusion chromatography (SEC). rMspA eluted at a volume of 12 ml buffer A. Fractions 12 and 13 were pooled. (D) Gel electrophoretic analysis of the purification steps. Proteins were separated on a denaturing 10% polyacrylamide gel. The gel was stained with Coomassie Blue. Lane M, molecular mass marker. For each lane of the gel, 3 μg protein was loaded. Lanes 1, crude cell lysate; 2, sample after AEC; 3, sample after HIC; 4, sample of lane 3 after precipitation with ammonium sulfate; 5, sample after SEC; 6, sample of lane 5 after ultrafiltration.

4B). The sample contained 15.9 mg protein of mainly rMspA and a few minor proteins with relative molecular masses exceeding 50 000, which were barely visible in a Commassie Blue-stained polyacrylamide gel (Fig. 4D, lane 3). To separate rMspA from the contaminating proteins and to remove the ammonium sulfate, the sample was first concentrated by ammonium sulfate precipitation and then purified further by size-exclusion chromatography (SEC) (Fig. 4C). It should be noted that ammonium sulfate precipitation induced the formation of the MspA tetramer with an apparent relative molecular mass of about 100 000 indicating that the purified rMspA monomer was folded correctly for association to the

channel-forming tetramer (Fig. 4D, lane 4). The rMspA tetramer and the contaminating proteins were almost completely removed from the rMspA monomer by SEC (Fig. 4D, lane 5). This step, however, resulted also in a loss of about 50% rMspA as evidenced by the drop in the amount of total protein from 15.9 to 7.3 mg. Finally, the protein was concentrated about 10-fold by ultrafiltration yielding 5.2 mg pure and water-soluble rMspA monomer with a concentration of 19.9 mg/ml (Fig. 4D, lane 6).

3.3. Analysis of purified rMspA

The rMspA monomer was recognized in immunoblot experiments by the anti-porin antiserum No. 813 [3] obtained with the native MspA protein from *M. smegmatis* (Fig. 5A). Surprisingly, the relative amount of tetrameric rMspA appeared to be higher than that of the rMspA monomer, although the rMspA tetramer was barely detectable in Commassie Blue-stained polyacrylamide gels (Fig. 4D). Increasing amounts of the tetrameric native MspA and the monomeric rMspA were blotted onto a nitrocellulose membrane to compare the binding of the anti-porin antiserum No. 813 to both proteins. The minimal amount of monomeric rMspA detected by the anti-porin antiserum in combination with chemoluminescence detection system was 10 ng (Fig. 5A, lane 7), whereas already 0.1 ng native tetrameric MspA gave a clear signal in the Western blot (Fig. 5B, lane 1). It appears unlikely that the M_r 20 000 monomer was less efficiently transferred to the membrane than the tetramer with an apparent relative molecular mass of 100 000, since it is known that the blotting efficiency decreases with the size of the protein [20]. Thus, it was concluded that detection of tetrameric MspA with the anti-porin antiserum No. 813 [3] was about 100-fold more sensitive than that of monomeric rMspA. Since the intensities of the bands of monomeric and tetrameric MspA were similar, it was estimated that the purified rMspA monomer contained about 1% tetrameric MspA (Fig. 5A, lane 7). This suggested that many epitopes recognized by the antiserum were only present in the MspA tetramer. The antiserum detected two oligomeric forms of rMspA similar to those observed after purification of native MspA from *M. smegmatis* [8].

The assembling of the rMspA monomer to the tetrameric MspA channel after ammonium sulfate precipitation (Fig. 4D, lane 4) suggested that at least a part of the monomer produced in *E. coli* was folded. This was analyzed by circular dichroism spectroscopy. The spectrum of the purified rMspA monomer showed a large positive mean molar ellipticity below 205 nm, a crossover point at 207 nm, at which the ellipticity equalled zero, and a single minimum at 215.5 nm (Fig. 6, solid line). These features are typical of β-sheet structures [21]

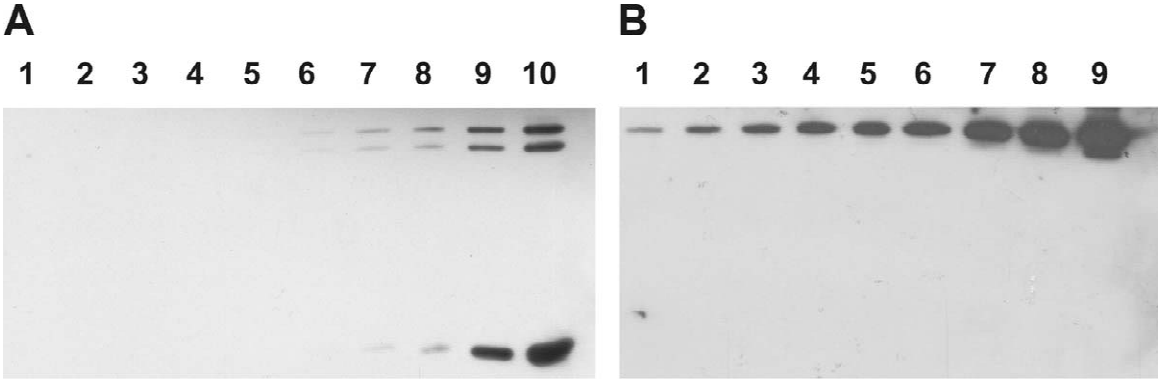

Fig. 5. Immunoblot analysis of rMspA. The samples were separated on a denaturing 10% polyacrylamide gel and blotted on a nitrocellulose membrane. Proteins were visualized using an MspA antiserum [3] and a chemoluminescence reaction (ECLplus detection system, Amersham Pharmacia, Vienna, Austria). (A) Increasing amounts of rMspA. Lanes: 1, 0.5 ng; 2, 1 ng; 3, 1.5 ng; 4, 2 ng; 5, 2.5 ng; 6, 5 ng; 7, 10 ng; 8, 15 ng; 9, 20 ng; 10, 40 ng. (B) Increasing amounts of MspA. Lanes: 1, 0.1 ng; 2, 0.2 ng; 3, 0.4 ng; 4, 0.6 ng; 5, 0.8 ng; 6, 1 ng; 7, 2 ng; 8, 5 ng; 9, 10 ng.

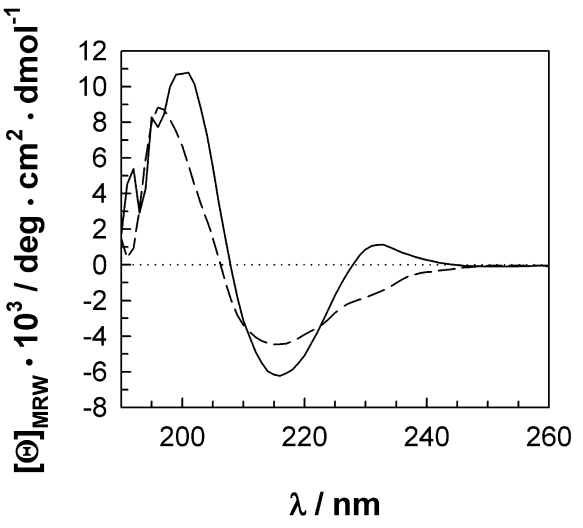

Fig. 6. Circular dichroism spectrum of rMspA. The mean residue molar ellipticity of purified rMspA (260 μg/ml, solid line) was recorded at 22 °C in a 25 m*M* phosphate buffer (pH 7.5). The spectrum of native MspA (260 μg/ml, dashed line) was recorded at 22 °C in a 25 m*M* phosphate buffer (pH 7.5) containing 0.5% *n*-octylpolyethylene oxide and was taken from Ref. [7]. Each spectrum was corrected for the buffer blank.

and were similar to those obtained for β-barrel proteins such as the porins of Gram-negative bacteria [22]. This demonstrated that the purified rMspA protein was folded. The positions of the crossover point and of the single minimum of the CD spectra of both the rMspA and the native MspA protein were almost identical (Fig. 6). However, the CD spectrum of the rMspA monomer did not reveal any α-helix content, in contrast to the tetrameric native MspA, which had an α-helix content of 10–15% [7].

4. Discussion

Identification and analysis of mycobacterial outer membrane proteins in general and of porins in particular have been hampered by the scarce amount of available protein. In this study, we present a straightforward approach to overcome this limitation. The *mspA* gene encoding for the major porin of *M. smegmatis* was synthesized with codons optimal for high-level expression in *E. coli*. When the *synmspA* gene was used, MspA was the most abundant protein in *E. coli*, whereas the mycobacterial *mspA* gene did not lead to detectable expression. This result demonstrated that high level expression of mycobacterial genes can be achieved in *E. coli*, when low-usage codons are replaced. Most likely this is due to a more efficient translation as was shown for three other mycobacterial genes [12], although we did not prove this formally by analysing the amount and stability of the mRNA transcribed from the *mspA* and *synmspA* genes. To increase expression of heterologous genes in *E. coli* often replacement of several rare codons in a single gene is necessary. This is feasible, but rather tedious when site-directed mutagenesis is used [12]. We showed that in vitro synthesis of a gene by assembly PCR from oligonucleotides [14] is conveniently achieved in a single day. Furthermore, optimal codons can be chosen for the entire gene, which might be of advantage for the quality of the produced protein, since a direct relationship has been established between the codon usage of a gene, the concentration of cognate tRNA species, the translation rate and the frequency of misincorporated amino acids [23–27].

Monomeric rMspA was purified in three chromatographic steps. The amount of protein of the cell lysate was reduced about eightfold after anion-exchange chromatography (Fig. 4). This indicated that many of the contaminating proteins were removed, since rMspA eluted in a sharp peak. Hydrophobic interaction chromatography was a very selective purification step, since rMspA bound more strongly than any other contaminating protein to the phenyl ether ligands of the column. Size exclusion chromatography was mainly used to remove small contaminating compounds such as ammonium sulfate from purified rMspA. The final yield was 2.6 mg rMspA per liter of culture. Thus, overexpression in *E. coli* and purification using the procedure described above increased the yield of MspA protein more than 10-fold compared to purification of MspA from *M. smegmatis* [8].

We showed by CD spectroscopy that purified rMspA was folded and consisted mainly of β-sheets similar to native MspA [7]. Furthermore, the salt concentration required for elution of rMspA was similar to that of the native tetrameric protein (510 vs. 570 m*M* NaCl) [8] indicating that a similar

number of negatively charged amino acids were surface-exposed in both proteins. The low ammonium sulfate concentration required for strong hydrophobic interaction with the HIC column indicated that the protein surface of rMspA was hydrophobic. Aggregation of rMspA was observed when a solution with a concentration of about 1 mg rMspA per ml became turbid in the presence of low amounts of ammonium sulfate after increasing the temperature to 37 °C. This effect was not observed without ammonium sulfate and was reversible upon cooling the sample on ice for a few minutes indicating that aggregation of rMspA was indeed due to hydrophobic interactions. Despite this apparent surface hydrophobicity, rMspA was soluble in water without detergents. A similar effect was observed for tetrameric MspA, which formed protein micelles and did not precipitate after removal of detergents by dialysis [5] and might indicate that also the solubility of monomeric rMspA in water is mediated by micelle formation.

About 15% of the recombinant MspA was present as tetramer after expression of the *synmspA* gene in *E. coli* (Fig. 3A), although the amount of rMspA tetramer varied with experimental conditions used for cell lysis. The recombinant tetramer was previously purified by preparative gel electrophoresis and was shown to form channels in lipid bilayer experiments [3]. This result demonstrated that the rMspA monomer was in the correct conformation to assemble to the functional MspA tetramer. The relative amount of the tetramer was about 1% after purification of the rMspA monomer as estimated from immunoblot experiments (Fig. 5). An increased amount of MspA tetramer was observed after ammonium sulfate precipitation suggesting that a high local protein concentration favoured formation of the tetramer. This might indicate that a small fraction of the rMspA monomer spontaneously assembled after purification. Based on these results, experiments are under way to improve the yield of the assembling reaction. Taken together, these data show that association of the recombinant MspA monomer to the functional channel tetramer is feasible and, therefore, provide for the first time the opportunity to produce quantities of a mycobacterial porin sufficient for X-ray crystallography and NMR spectroscopy. A high resolution structure of a mycobacterial porin would be of paramount importance for understanding matter exchange across the mycobacterial cell wall.

Acknowledgements

We thank Dr Wolfgang Hillen for generous support. This work was funded by the Deutsche Forschungsgemeinschaft (NI 412) and the Volkswagen-Stiftung (I/77 729).

References

[1] P.J. Brennan, H. Nikaido, Annu. Rev. Biochem. 64 (1995) 29.
[2] J. Trias, V. Jarlier, R. Benz, Science 258 (1992) 1479.
[3] M. Niederweis, S. Ehrt, C. Heinz, U. Klöcker, S. Karosi, K.M. Swiderek, L.W. Riley, R. Benz, Mol. Microbiol. 33 (1999) 933.
[4] C. Stahl, S. Kubetzko, I. Kaps, S. Seeber, H. Engelhardt, M. Niederweis, Mol. Microbiol. 40 (2001) 451.
[5] H. Engelhardt, C. Heinz, M. Niederweis, J. Biol. Chem. 277 (2002) 37567.
[6] G.E. Schulz, Curr. Opin. Struct. Biol. 6 (1996) 485.
[7] C. Heinz, H. Engelhardt, M. Niederweis, J. Biol. Chem. (2003) in press.
[8] C. Heinz, M. Niederweis, Anal. Biochem. 285 (2000) 113.
[9] N.D. Connell, Curr. Opin. Biotechnol. 12 (2001) 446.
[10] F.R. Blattner, G. Plunkett 3rd, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, Y. Shao, Science 277 (1997) 1453.
[11] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry 3rd, 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, B.G. Barrell, Nature 393 (1998) 537.
[12] D.L. Lakey, R.K. Voladri, K.M. Edwards, C. Hager, B. Samten, R.S. Wallis, P.F. Barnes, D.S. Kernodle, Infect. Immun. 68 (2000) 233.
[13] F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidmann, J.A. Smith, K. Struhl, Current Protocols in Molecular Biology, Wiley, New York, 1987.
[14] W.P. Stemmer, A. Crameri, K.D. Ha, T.M. Brennan, H.L. Heyneker, Gene 164 (1995) 49.
[15] P.K. Smith, I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Anal. Biochem. 150 (1985) 76.
[16] F.W. Studier, B.A. Moffatt, J. Mol. Biol. 189 (1986) 113.
[17] F. Baneyx, Curr. Opin. Biotechnol. 10 (1999) 411.
[18] A. Hénaut, A. Danchin, in: F.C. Neidhardt (Ed.), *Escherichia*

coli and *Salmonella*, ASM Press, Washington, DC, 1996, p. 2047.
[19] F. Vilbois, P. Caspers, M. da Prada, G. Lang, C. Karrer, H.W. Lahm, A.M. Cesura, Eur. J. Biochem. 222 (1994) 377.
[20] W. Gibson, Anal. Biochem. 118 (1981) 1.
[21] N. Greenfield, G.D. Fasman, Biochemistry 8 (1969) 4108.
[22] M. Schindler, J.P. Rosenbusch, FEBS Lett. 173 (1984) 85.
[23] O.G. Berg, C.G. Kurland, J. Mol. Biol. 270 (1997) 544.
[24] M.A. Sorensen, C.G. Kurland, S. Pedersen, J. Mol. Biol. 207 (1989) 365.
[25] K. Zahn, J. Bacteriol. 178 (1996) 2926.
[26] M.D. Forman, R.F. Stack, P.S. Masters, C.R. Hauer, S.M. Baxter, Protein Sci. 7 (1998) 500.
[27] S. Zhang, E. Goldman, G. Zubay, J. Theor. Biol. 170 (1994) 339.