Volume 292, number 1,2, 90–94 FEBS 10326 November 1991

ADONIS 0014579391010368

The Ssc protein of enteric bacteria has significant homology to the acyltransferase Lpxa of lipid A biosynthesis, and to three acetyltransferases

Riitta Vuorio, Laura Hirvas and Martti Vaara

Department of Bacteriology and Immunology, University of Helsinki, Haartmaninkatu 3, 00290 Helsinki, Finland

Received 12 August 1991; revised version received 10 September 1991

The Ssc protein, a novel essential protein affecting the function of the enterobacterial outer membrane, matched in a protein homology search best with LpxA (UDP-*N*-acetylglucosamine 3-hydroxymyristoyl transferase), the enzyme which catalyzes the first step of lipid A biosynthesis. The corresponding genes, located 0.56 kb apart, were 46.7% identical. The search also revealed homology to the bacterial acetyltransferases LacA and NodL, as well as to a hypothetical protein Yglm. The region of residues 109–149 Ssc displayed the highest homology and was also homologous with another bacterial acetyltransferase, CysE, and three other bacterial proteins, two of which are hypothetical. This region and the corresponding regions of all other proteins were found to have a peculiar repeated hexapeptide pattern. Each hexapeptide unit starts with isoleucine (or its equivalent leucine and valine). In most units, the second residue is glycine and the fifth residue either valine or alanine.

Ssc; LpxA; Lipid A biosynthesis; Acetyltransferase; *Enterobacteriaceae*

1. INTRODUCTION

The outer membrane (OM) of Gram-negative bacteria is a complex biological membrane and an effective permeability barrier, especially against hydrophobic solutes such as hydrophobic antibiotics [1]. While the most conserved part of the lipopolysaccharide (LPS) constituent of the OM (i.e. the lipid A-KDO-heptose part) is essentially involved in the permeability barrier function, the other specific components of the OM, including the remaining approx. 90% of LPS (the outer core and O-serologic part), are not as essential [1]. Notably, defects in the biosynthesis of the lipid A-KDO region are lethal while those in the biosynthesis of the other parts of LPS are not [2]. This has greatly hampered the isolation of lipid A mutants and, as a consequence, only two of the more than five enzymes involved in lipid A biosynthesis have been characterized in molecular terms.

We have recently discovered a novel essential enterobacterial gene which affects the function of the OM [3,4]. This gene, *ssc*, encodes a 35.9 kDa protein and is located in the bacterial chromosome between *ompH* (the gene which encodes the cationic 17 kDa outer membrane protein) and the known two genes (*lpxA*, *lpxB*) involved in lipid A biosynthesis. The *Salmonella typhimurium* mutant SH7622, which carries the mutant allele *ssc1*, is extremely sensitive to numerous hydrophobic antibiotics at growth-permissive temperatures and does not grow at 42°C [4,5]. *S. typhimurium ssc* and its *Escherichia coli* analogue *firA* are highly homologous (identity, 88%) as are the corresponding proteins Ssc and FirA (identity, 96%) [4,6]. We are currently trying to identify the function of Ssc. It could be plausible to think that it is involved in lipid A biosynthesis.

In this communication, we show that Ssc shares partial but significant homology to the LpxA protein (UDP-acetylglucosamine acyltransferase), which catalyzes the first step of lipid A biosynthesis. This homology is also clearly seen at the nucleotide level. In addition, Ssc is partially homologous with three acetyltransferases as well as with succinyldiaminopimelate aminotransferase and three hypothetical proteins. Furthermore, we will show that Ssc, LpxA and all these other proteins possess a 41 amino acid-long highly conserved region which has a repeated hexapeptide pattern and a consensus sequence.

Abbreviations: bp, base pair(s); KDO, ketodeoxyoctulosonic acid; LPS, lipopolysaccharide; OM, outer membrane; ORF, open reading frame

Correspondence address: M. Vaara, Department of Bacteriology and Immunology, University of Helsinki, Haartmaninkatu 3, 00290 Helsinki, Finland. Fax: (358) (0) 434 6382.

2. MATERIALS AND METHODS

If not otherwise stated, the protein homology searches were performed by using the FASTA program [7] (*ktup* 1) of the University of Wisconsin Genetics Computer Group program package (GCG [8], version 6.2, June 1990) and the SWISS-PROT protein sequence library (European Molecular Biology Laboratory, release 15, November 1990). The homology in the Ssc region extending from the amino acid

Published by Elsevier Science Publishers B.V.

residue 109 to the residue 149 was studied using the HIBIO PROSIS program package (Hitachi Software Engineering Co., Ltd., Yokohama, Japan, *ktup* 1) and the SWISS-PROT database (release 14, April 1990) included in this package.

In homology studies, the equivalent amino acids were as follows: A, G, S, T, P; E, D, N, Q; H, K, R; I, L, M, V; F, Y, W; C. The diagonal homology plots were made by using the COMPARE and DOTPLOT programs (in GCG) with the stringency of 60% and the window of 25 residues.

The DNA homology was studied by using the HIBIO DNASIS program (Hitachi, *ktup* 4) and the GenBank database (National Institutes of Health, USA; release 64, June 1990).

3. RESULTS

3.1. *Ssc has significant homology with LpxA*

The homology of Ssc with other proteins was studied by employing the SWISS-PROT database and the FASTA search program. From all database proteins, best match was found with *E. coli* UDP-acetylglucosamine acyltransferase (LpxA), a component of the biosynthesis of lipid A [9].

The alignment of the Ssc protein and the LpxA protein is shown in Fig. 1. When hypothetical deletions are omitted, a 223 amino acid-long region with the identity of 22.4% (50 identical amino acids) can be found. If the 48 additional 'equivalent' amino acids (specified as in section 2) are included, the homology is 44%.

The homology was expectedly noted also at the DNA level. When the structural *ssc* gene of *S. typhimurium* was compared with all GenBank bacterial genes by using the DNASIS program, the best homology (after optimization) was obtained with *lpxA* of *E. coli* (46.7% identity in a 707 bp overlap with the structural *lpxA* gene).

3.2. *The homology of Ssc with other proteins*

FASTA revealed, in addition to LpxA, three other proteins with an initial homology score of more than six SDs higher than the mean score (32;1 SD, 6.7). Those were *E. coli* thiogalactoside acetyltransferase (LacA; 91, 104), *E. coli* Yglm (the urf1 protein; 83, 137), and the *Rhizobium leguminasorum* nodulation protein L (NodL; 80, 91) (values in parentheses are the initial and optimized scores, respectively; the correspondings scores for LpxA were 98 and 176).

The acetyltransferase LacA is known to share homology with another *E. coli* acetyltransferase, serine acetyltransferase (CysE) [10]. The NodL protein is probably also an acetyltransferase and has homology with LacA and CysE [10]. Yglm is a 49.2 kDa hypothetical protein encoded by the flanking region of the *glmS* (glucosaminephosphate isomerase) gene [11].

Fig. 2 shows the comparison of Ssc with LpxA, Yglm,

```
Ssc    MPSIRLADLAEQLDAELHGDGDIVITGVASMQSATTGHITFMVNPKYREHLGLCQASAVVMTQDDLPFAK-  (70)

Ssc    SAALVVKNPYLTYARMAQILDTTPQPAQNIAPSAVIDATATLGSNVSVGANAVIESGVQLGDNVVIGAGC- (140)
                                     |  || : :|| :    |:|||| | :   :|  | || |
LpxA                                MIDKSAFVHPTAIVEEGASIGANAHIGPFCIVGPHVEIGEGT-  (42)

Ssc    FVGKNSKIGAGSRLWANVTIYHDIQIGENCLIQSSTVIGADGFGYANDRGNWVKIPQLGRVIIGDRVEIG- (210)
        :    : : ::: :   ||    |||       :                    : :|| |||| |
LpxA   VLKSHVVVNGHTKIIGRDEIYSVASIGEVNQDLKYAG-------------------EPTRVEIGDRNRIR-  (93)

Ssc    ACTTIDRGALDDTV---IGNGVIIDNQCQIAHNVVIGDNTAVAGGVIMAGSLKIGRYCMIGGASVINGHM- (277)
          || ||::: :    :|   ::  :  || :  :|:   :|    :|| : :  : :||| : :
LpxA   ESVTIHRGTVQGGGLTKVGSDNLLMINAHIADDCTVGNRCILANNATLAGHVSVDDFAIIGGMTAVHQFC- (163)

Ssc    EICDKVTVTGMGMVMRPITEPGVYSSGIPLQPNKVWRKTAALVMNIDDMSKRLKAIERKVNQQD         (341)
        |  :| |:| : |   ::   : : |    |  |   :        :    ::   : :
LpxA   IIGAHVMVGGCSGVAQDVPPYVI-AQGNHATPFGVNIEGLKRRGFSREAITAIRNAYKLIYRSGKTLDEV- (232)

LpxA   KPEIAELAETYPEVKAFTDFFARSTRGLIR                                          (262)
```

Fig. 1. Alignment of the Ssc protein of *S. typhimurium* and the LpxA protein (UDP-*N*-acetylglucosamine 3-hydroxymyristoyl transferase) of *E. coli*. Besides identical amino acids (marked with a stripe), 'equivalent' amino acids (marked with a colon) are also shown.

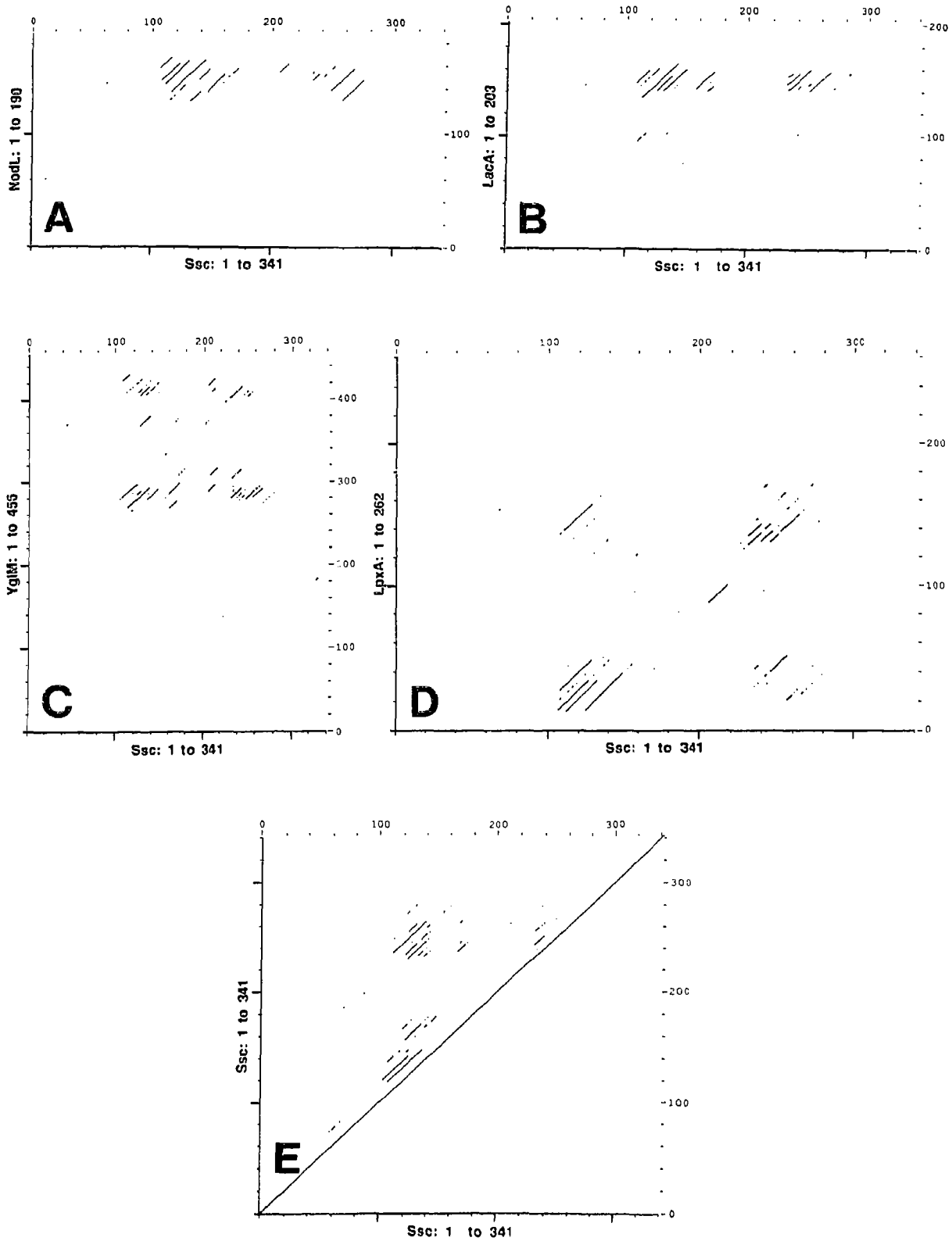

Fig. 2. Diagonal comparison of protein sequences. All comparisons were made with the proportional matrix using the COMPARE the DOTPLOT programs of the GCG package (See section 2). Ssc was compared with the following proteins: A, NodL; B, LacA; C, YglM; D, LpxA. Panel E is the Ssc self comparison.

LacA, and NodL by using diagonal homology plots. The region of Ssc between residues approx. 100–180 (region 100–180) was homologous with certain regions of all these proteins. Another region of Ssc which dis-

```
FirA/E.coli   (109)   TAK LGNNVS IGANAV IESGVE LGDNVI IGAGCF VGKNSK IG    (149)
Ssc/S.tm      (109)   ♦♦T ♦♦S♦♦♦ V♦♦♦♦♦ ♦♦♦♦♦Q ♦♦♦♦♦V ♦♦♦♦♦♦ ♦♦♦♦♦♦ ♦♦    (149)
LpxA/E.coli    (11)   ♦♦I VEEGA♦ ♦♦♦♦♦H ♦GPFCI V♦PH♦E ♦♦E♦TV LKSHVV VN    (51)
LacA/E.coli   (131)   PIT I♦♦♦♦W ♦♦SHV♦ ♦NP♦♦T I♦♦♦SV ♦♦♦♦SI ♦T♦DIP PN    (171)
NodL/R.leg.   (131)   PVS I♦RHAW ♦♦GG♦♦ ♦LP♦♦T I♦♦HAV ♦♦♦♦SV ♦TRDVP A♦    (171)
CysE/E.coli   (193)   HP♦ IREG♦M ♦♦♦G♦K ♦LGNI♦ V♦RGAK ♦♦♦♦SV ♦LQPVP PH    (233)
Yglm/E.coli   (265)   ♦LT H♦RD♦E ♦DT♦VI ♦♦GN♦T ♦♦HR♦K ♦♦T♦♦V I-♦♦♦V ♦♦    (304)
Tms/B.subt.   (266)   D♦V I♦SDTV ♦YPGT♦ ♦KGE♦Q I♦EDT♦ ♦♦PHTE I-M♦♦A ♦♦    (305)
DapD/E.coli   (175)   PTM IED♦CF ♦♦♦RSE LVE♦♦I VEEGSV ♦SM♦VV I♦QSTR ♦Y    (195)
Yerm/B.sph.   ( ? )   DTV I♦♦D♦W ♦♦Q♦VT ♦MP♦♦I I♦♦GA♦ ♦A♦NST ♦V♦SVE PY    ( ? )

consensus             T-- IG--V- IG--A- I---V- IG--VV IG-G-- V------ --
                      P   L   A  V   V  L      V   AI       I
                          V                    L            L

                      7-- 97--8- 18--8- 1---8- 19--77 18-8-- 1----- --
                                 0       0       0       0       0
```

Fig. 3. Alignment of the region of amino acid residues 109–149 of the Ssc protein (and its analogue FirA) with the most homologous parts of eight other proteins. The black diamonds (♦) indicate amino acids identical to those of Ssc. Amino acids not identical but equivalent are underlined with a double line. The deletions are marked with –. A partial consensus sequence (provided with several alternative amino acid residues) is also given. Vertically, beneath the consensus, the number (7 through 10) of matches per position is given.

played homology was the region 200–280. Furthermore, Ssc–Ssc self-comparison (Fig. 2E) indicated that these two regions of Ssc were homologous with each other, too. Fig. 2E also reveals the existence of multiple very short homologous repeated sequences within the Ssc region approx. 100–160.

We then started to zoom in on the region of highest homology and performed a FASTA search to find homologies with the region 100–200 of the Ssc protein. Among all proteins, the best scores were again obtained with Lpxa, LacA, Yglm, and NodL, in this order. The same proteins had best scores in a search for homology with the region 100–150. The latter search also revealed two additional bacterial proteins with significant homology, succinyldiamino-pimelate aminotransferase (DapD) of *E. coli* and the Tms protein of *Bacillus subtilis*. Tms is a hypothetical protein which has strong homology with the Yglm protein of *E. coli* [12].

A DNA homology search by DNASIS indicated that, next to *lpxA*, the highest optimized scores were achieved with the *lac* operon of *e. coli* (homology with the structural *lacA* gene) and the *unc* operon of *E. coli* (homology with the ORF encoding Yglm). This search also revealed homology with the gene region *ermG* [13] of *Bacillus sphaericus* (homology with the ORF immediately upstream of *ermG*, no deduced amino acid sequence being deposited in SWISS-PROT). The hypothetical peptide, encoded by this ORF and called here as Yerm shared notable homology with Ssc and again, this homology centered on the region 100–200 of Ssc. Another protein which has been reported to share homology with Yerm is NodL [10].

3.3. *A common hexapeptide repeat motif*

In Fig. 3, the region 109–149 of the Ssc/FirA protein has been aligned, by using the PROSIS program, with the most homologous parts of all the above-mentioned actual or hypothetical proteins (LpxA, LacA, CysE, NodL, Yglm, Tms, DapD, Yerm). The homology of all eight sequences with the FirA sequence is remarkable (homology score range, 54–87; mean score, 68; SD, 12). The lowest scores were obtained with CysE and Tms and differed from the databank mean score (mean, 22; SD, 5.7) by 5.5 standard deviations. The alignment reveals a partial but long consensus sequence. Even more strikingly, a peculiar six-residue periodicity can be found in the consensus. Each of the six hexapeptide units starts with isoleucine, leucine, or valine. In four of the units, the second residue is glycine. Alanine or valine is the fifth residue in four units.

4. DISCUSSION

In this paper we showed that the most homologous protein to Ssc is the acyltransferase LpxA, the enzyme which catalyzes the transfer of β-hydroxymyristic acid from the acyl carrier protein (ACP) to the 3-hydroxy position of UDP-*N*-acetylglucosamine. In a 223 amino acid-long common region, the identity was 23.3% (and the homology allowing equivalent amino acids, 42.6%).

Consistently, among all Genbank bacterial DNA sequences, the most homologous sequence to *ssc* was found to be within the structural *lpxA* gene. Furthermore, we have recently found that the conditionally lethal *lpxA2* mutant [14] is at growth-permitting temperatures extremely sensitive to hydrophobic antibiotics (R. Vuorio and M. Vaara, submitted for publication), as is the *ssc1* mutant [4].

Accordingly, it would be tempting to speculate that Ssc is involved in lipid A biosynthesis and that it has a function which resembles that of LpxA, i.e. an acyltransferase function. Studies with bacterial cell fractions indicate that, besides LpxA, at least three other acyltransferases participate in lipid A biosynthesis [2]. They are UDP-3-monoacyl-*N*-glucosamine β-hydroxymyristoyltransferase as well as the 'late' acyltransferases, namely KDO_2-lipidIV_A lauroyl- and myristoyltransferases. None of them have been characterized at a protein or DNA level as yet, but all are known to transfer the fatty acid to the lipid A precursor from the fatty acyl-ACP thioester as does LpxA. We are going to test whether our thermosensitive *ssc1* mutant is defective in any of those enzyme activities at 42°C.

Our searches also revealed homology with three acetyltransferases (LacA, CysE, and NodL) and, to a lesser extent, with succinyldiaminopimelate aminotransferase and three hypothetical proteins. We found that all these proteins possess a 41 amino acid-long highly conserved region. An earlier report mentioned that the *E. coli* analogue, FirA, has homology with LacA and CysE (as well as to LpxA), but provided no data [6]. LacA, CysE and NodL have previously been shown to be remarkably homologous with each other and to display a common I-G-A-G-S-[L,I,V,M]-V motif, named as the 'the acetyltransferase motif'. On this basis, these enzymes have been suggested to form a single family of bacterial acetyltransferases [10] (see also the PROSITE database [15]). The acetyltransferase motif can also be found in Fig. 3 (residues in positions 28–34 of the LacA, CysE, and NodL alignment, starting at the first residue of the fifth hexapeptide unit). The corresponding residues of Ssc/Fira, LpxA, and Yglm alignments were remarkably homologous and those of Tms, DapD, and yerm partially homologous with the acetyltransferase motif. The significance of the homology to these acetyltransferases remains open. No homology was found with the other acetyltransferases. Both LacA and CysE utilize acetyl-CoA and thus differ from lipid A acyltransferases which use fatty acyl-ACP (not fatty acyl-CoA).

The 41 amino acid-long region of homology has certain unique and interesting characteristics. It has a peculiar six-residue periodicity. Each hexapeptide unit starts with isoleucine (or its equivalents, leucine and valine). In most units, the second residue is glycine and the fifth residue alanine or valine. This pattern favors β-turns (as calculated by the method of Chou and Fasman [16]) and is distantly reminiscent to the tripeptide periodicity in α collagens [17] and the tri- and pentapeptide periodicity in elastins [18].

Acknowledgements: This work was supported by grants 01/690 and 1011749 from the Academy of Finland (to M.V.) and by the Sigrid Juselius Foundation (to M.V. and R.V.).

REFERENCES

[1] Nikaido, H. and Vaara, M. (1985) Microbiol. Rev. 49, 1–32.
[2] Raetz, C.R.H. (1990) Annu. Rev. Biochem. 59, 129–170.
[3] Hirvas, L., Koski, P. and Vaara, M. (1990) Biochem. Biophys. Res. Commun. 173, 53–59.
[4] Hirvas, L., Koski, P. and Vaara, M. (1991) EMBO J. 10, 1017–1023.
[5] Vaara, M., Hirvas, L. and Koski, P. (1990) in: Cellular and Molecular Aspects of Endotoxin Reactions (Nowotny, A., Spitzer, J. and ziegler, E., eds.) pp. 197–204, Elsevier, Amsterdam.
[6] Dicker, I. and Seetharam, S. (1991) J. Bacteriol. 173, 334–344.
[7] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[8] Devereaux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 389–391.
[9] Coleman, J. and Raetz, C.R.H. (1988) J. Bacteriol. 170, 1268–1274.
[10] Downie, J.A. (1989) Mol. Microbiol. 3, 1649–1651.
[11] Walker, J.E., Gay, N.J., Saraste, M. and Eberle, A.N. (1984) Biochem. J. 224, 799–815.
[12] Nilsson, D., Hove-Jensen, B. and Arnvig, K. (1989) Mol. Gen. Genet. 218, 565–571.
[13] Monod, M., Mohan, S. and Dubnau, D. (1987) J. Bacteriol. 169, 340–350.
[14] Galloway, S.M. and Raetz, C.R.H. (1990) J. Biol. Chem. 265, 6394–6402.
[15] Bairoch, A. (1990) A dictionary of protein sites and patterns, University of Geneva.
[16] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.
[17] Fietzek, P.P., Rexrodt, F.W., Hopper, K.E. and Kuehn, K. (1973) Eur. J. Biochem. 38, 396–400.
[18] Raju, K. and Anwar, R.A. (1987) J. Biol. Chem. 262, 5755–5762.