ELSEVIER

FEBS Letters 342 (1994) 23-28

FEBS 13820

FEBS LETTERS

Binding site-shaped repeated sequences of bacterial wall peptidoglycan hydrolases

Jean-Marie Ghuysen[a,*], Josette Lamotte-Brasseur[a], Bernard Joris[a], Gerald D. Shockman[b]

[a]*Centre d'Ingénierie des Protéines, Université de Liège, Institut de Chimie, B6, B-4000 Sart Tilman (Liège 1), Belgium*
[b]*Department of Microbiology and Immunology, Temple University School of Medicine, 3400 North Broad Street, Philadelphia, PA 19140, USA*

Received 7 February 1994

Abstract

The non-catalytic C-terminal regions of the *N*-acetylmuramidase (lysozyme) of *Clostridium acetobutylicum* and *N*-acetylmuramoyl(D-lactyl)-L-alanine amidases CwlA of *Bacillus subtilis*, ORFL3 and CwlL of *Bacillus licheniformis* were previously reported to have similarities with the amino acid sequence of the non-catalytic N-terminal module of the *Streptomyces albus* G Zn DD-peptidase. This peptidase is a bipartite protein of known three-dimensional structure. Its non-catalytic N-terminal module possesses, exposed at the surface, an elongated crevice which is defined by a loop-helix-loop-helix motif that consists of two repeats, each 16 amino acid residues long, connected by a heptapeptide and whose design is compatible with its possible functioning as a substrate recognition and binding site. Amino acid alignments suggest that cavities nearly identical in shape to that present in the non-catalytic module of the *S. albus* peptidase, are borne by the C-terminal regions of the CwlA amidase (in one copy), the lysozyme and the ORFL3 and CwlL amidases (in two copies). Since a common feature of the five enzymes is their substrate, the bacterial cell wall peptidoglycan, we interpret the striking similarity of their non-catalytic N- or C-terminal modules to suggest that these modules are involved in the binding of these exocellular enzymes to their insoluble wall substrate.

Key words: Protein modular design; Peptidoglycan hydrolase; Repeated sequence; 3-D Structure

1. Introduction

A number of bacterial wall lytic enzymes [1] and cell-surface-associated proteins of Gram-positive bacteria [2] appear to be constructed in a modular fashion. Whereas one domain contains the catalytic site, another domain, which is frequently at or near the N- or C-terminal end of the amino acid sequence, contains two or more tandemly repeated sequences, 20 or more amino acids long. There appears to be several families of repeated sequences and in more than one instance the repeats have been associated with ligand binding domains.

The most direct experimental evidence that such domains are involved in ligand binding was obtained by genetic engineering of pneumococcal bacteriolytic enzymes [3,4]. For example, the portion of the gene encoding the catalytic N-terminal domain of the pneumococcal choline-dependent LytA amidase [5] was fused to the portion of the gene encoding the non-catalytic C-terminal domain of the choline-independent lysozyme, LC7, of bacteriophage CP7 [6]. The resultant chimeric enzyme expressed in *Escherichia coli* possessed an amidase activity that was independent of the presence of choline. The reciprocal chimeric enzyme containing the N-terminal domain of the bacteriophage choline-independent lysozyme fused to the C-terminal domain of the pneumococcal choline-dependent amidase hydrolysed the lysozyme-sensitive bonds only in choline-containing cell walls. In spite of these advances, proof, in terms of three-dimensional structure, that the non-catalytic domains possess substrate recognition and binding sites is still lacking.

The goal of this report is to show that a potential binding site is present in the non-catalytic N-terminal module of the *Streptomyces albus* G metallo (Zn) DD-peptidase [7] and, by analogy, in the C-terminal regions of the *Clostridium acetobutylicum* lysozyme [8], *Bacillus subtilis* CwlA amidase [9,10], *Bacillus licheniformis* FD0120 CwlL amidase [11] and *B. licheniformis* MC14 ORFL3 gene product [12]. The function of the ORFL3 gene product has not been determined but its derived amino acid sequence is similar to that of the CwlL amidase [11], strongly suggesting that the ORFL3 gene product is also an amidase. Although frequently referred to as a muramoyl pentapeptide carboxypeptidase, the *S. albus* peptidase hydrolyses C-terminal D-alanyl-(D)-diamino acid linkages which serve as interpeptide bridges in the wall peptidoglycan of various bacterial species [13,14]. Hence, the *S. albus* peptidase, the *C. acetobutylicum* lysozyme and the three *Bacillus* amidases are peptidoglycan hydrolases of varying enzymatic specificities. They hydrolyse linkages within the peptide moiety, the

*Corresponding author. Fax: (32) (41) 56 33 64.

SSDI 0014-5793(94)00205-A

Fig. 1. Mono and stereo views of the folding of the α-carbon polypeptide chain of the *Streptomyces albus* G Zn DD-peptidase. The drawings and those of Figs. 3 and 4 were made by using the PLUTO programme on a Silicon Graphics workstation and the Insight programme of Biosym.

glycan chains and at the junction between the glycan and peptide moieties, respectively.

2. Results

The *S. albus* peptidase is a 213 amino acid residue bipartite protein. Its three-dimensional structure has been resolved to 0.18 nm [15,16]. Fig. 1 shows the folding of the α-carbon polypeptide. The catalytic, 132 amino acid residue C-terminal module possesses three α-helices (α4–α6) and five-stranded β-sheet. The open arrow on the right points toward the zinc-containing site. Fused to the catalytic module, the non-catalytic, 81 amino acid residue N-terminal module possesses three α-helices (α1–α3). The open arrow on the left points toward a cavity which is defined by the sequence L36–Q77. This sequence comprises the C-terminal portion of the loop that connects helix α1 to helix α2, helix α2, the loop that connects helix α2 to helix α3, and helix α3.

The amino acid sequence alignment of Fig. 2 highlights the repeated design of the sequence L36–Q77. Repeat 1 extends from L36 to Q54 and repeat 2 extends from L59 to Q77. As shown in Fig. 3, repeat 1, from D39 to Q54, and repeat 2, from D62 to Q77, have almost identical polypeptide foldings. Superimposition of the N-Cα-C = O groupings occurs within a range of 0.02 to 0.05 nm. The side chains of the pair D39 and D62, the pair P44 and P67, the pair A45 and A68, the pair T46 and T69, the pair V50 and I73 (up to the Cγ atoms) and

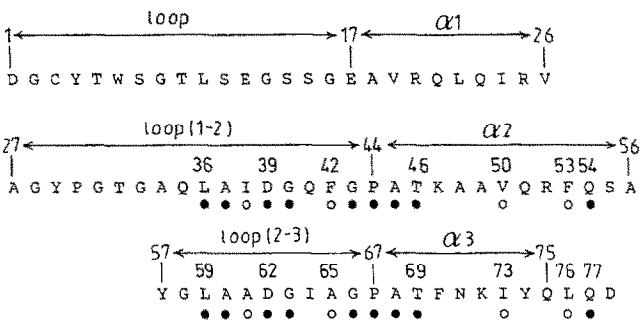

Fig. 2. Amino acid sequence of the N-terminal module of the *Streptomyces albus* G Zn DD-peptidase. The positions of the secondary structures are indicated. The alignment highlights the repeats L36–Q54 and L59–Q75. Filled circles identify strict identities. Open circles identify hydrophobic residues occuring at equivalent places.

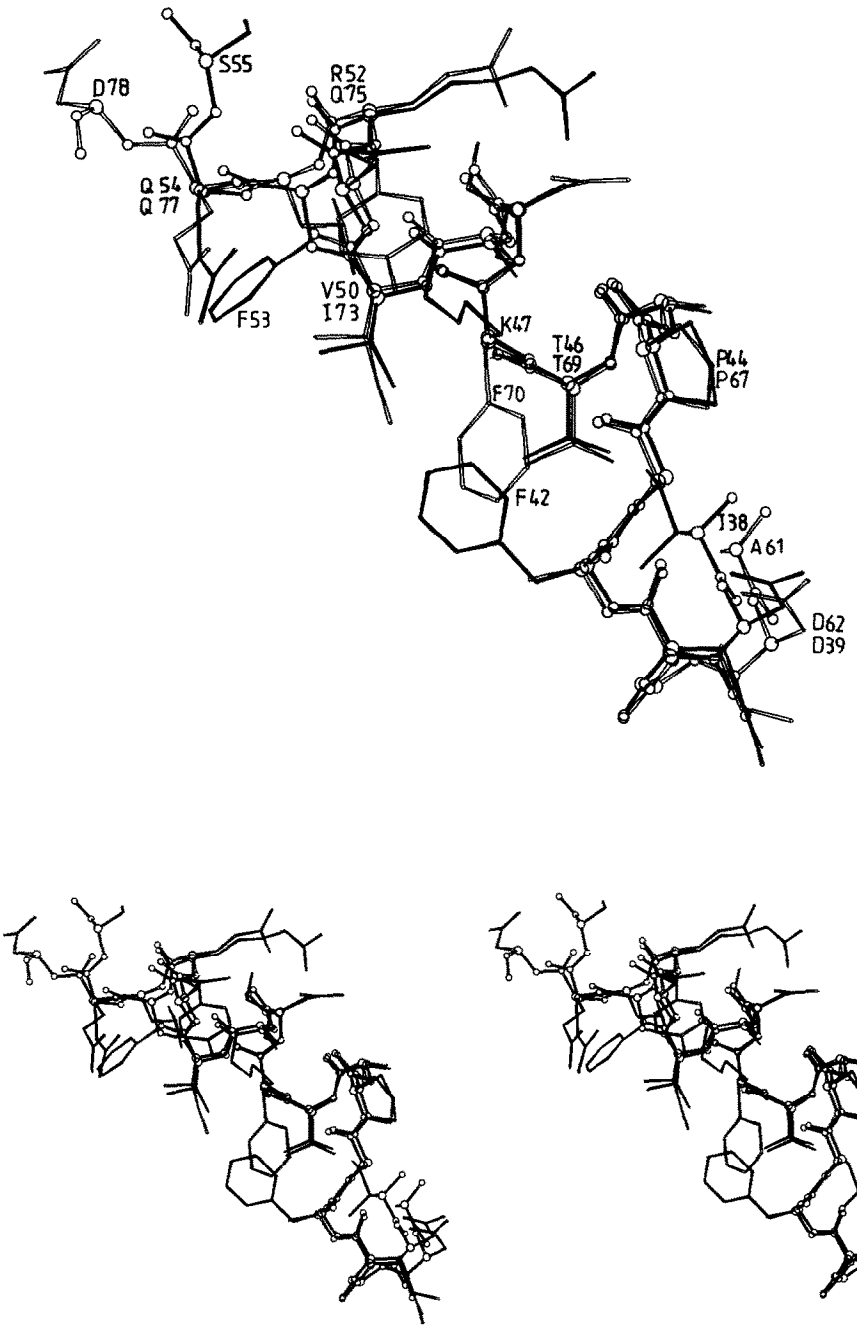

Fig. 3. Mono and stereo views of the superimposed polypeptides D39–S55 (filled lines) and D62–D78 (open lines) of the N-terminal module of the *Streptomyces albus* G Zn DD-peptidase.

the pair R52 and Q75 (up to the Cδ atoms) superimpose remarkably well. It thus follows that the cavity is defined by two superimposable repeats, D39–Q54 and D62–Q77 each 16 amino acid residues long, connected by the heptapeptide S55–A61.

Fig. 4 presents a view of the cavity. It also shows the spatial disposition of the side chains of D39, F42, T46, V50, F53 and Q54 of repeat 1, the side chains of D62, I64, T69, F70 and I73 of repeat 2 and the side chains of S55 and L59 of the intervening heptapeptide S55–A61. The side chains of D62 and I64 are at the left entrance of the cavity, the side chains of D39 and I73 are at the right entrance, the side chains of T46, T69 and Q54 and those of the hydrophobic amino acid residues F42, V50, F53 and L59 occupy central positions. The spanning distances between the Oδ atoms of D39 and D62 and between the Oγ atoms of T46 and T69 are 1.86 nm and 1.35 nm, respectively. The tetrapeptide Y74–Q77 at the C-terminal end of repeat 2 and the side chains of S55 and F70 are outside the cavity.

On the basis of these structural data, we have used the non-catalytic, N-terminal module of *S. albus* peptidase

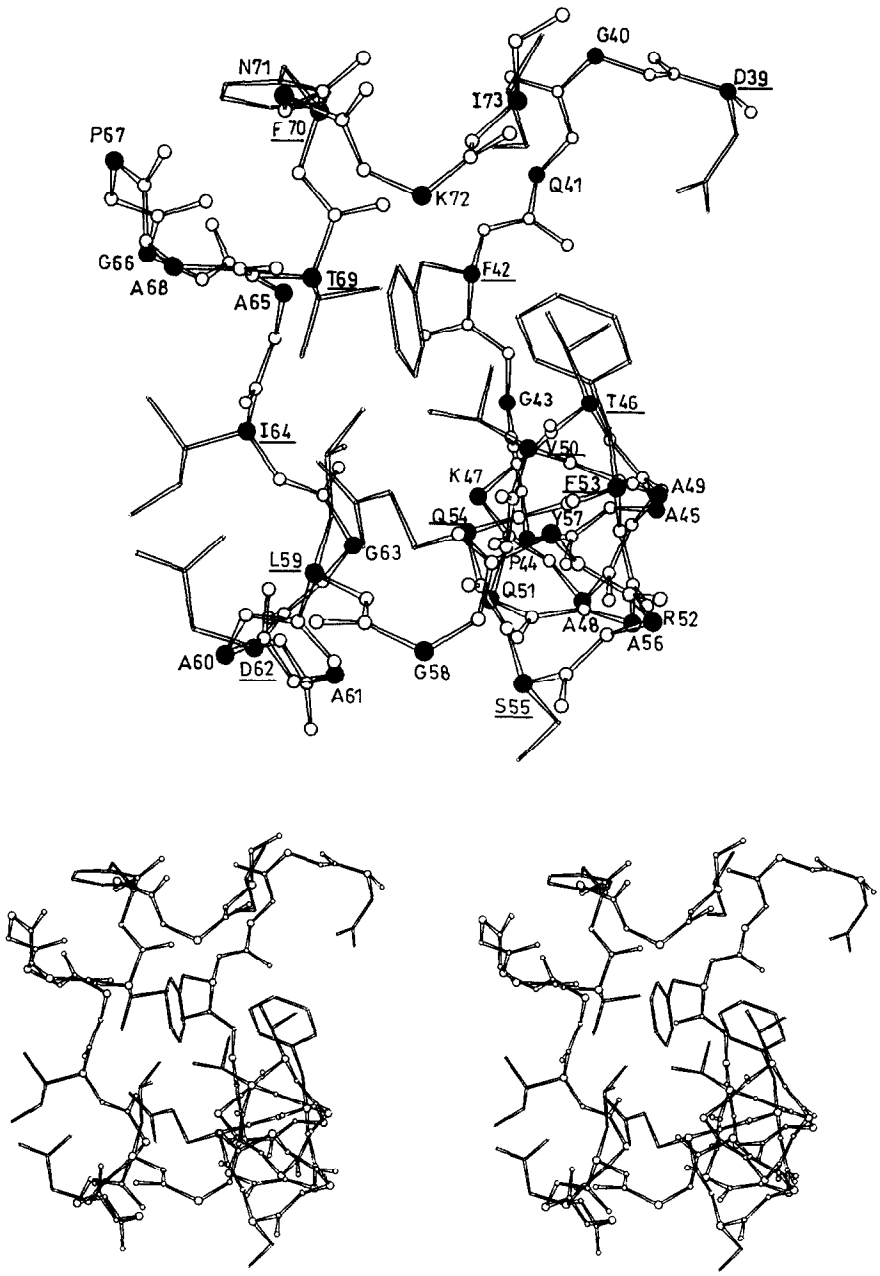

Fig. 4. Mono and stereo views of the crevice created by the polypeptide D39–I73 at the surface of the N-terminal module of the *Streptomyces albus* G Zn DD-peptidase. The α-carbon are filled circles. The N-CO atoms are open circles. The lateral chains of selected amino acid residues (underlined) are shown.

as a model in an attempt to shed light on the structural and therefore possible functional features of the non-catalytic C-terminal regions of the *C. acetobutylicum* lysozyme and *Bacillus* CwlA, CwlL and ORFL3 amidases. The amino acid alignments shown in Fig. 5 were made by using the best fit algorithm of the GCG package [17]. The following observations can be made.

(i) The C-terminal region of the CwlA amidase, from M204 to K272, has 50% identities with the N-terminal module of the *S. albus* peptidase, from S14 to Q77.

(ii) The C-terminal regions of the ORFL3 and CwlL amidases are of more complex design. Each is comprised of three polypeptide segments. Two, M199–E262 and M286–K354 in ORFL3, and A200–G263 and M292–K360 in CwlL, are, respectively, 64 and 69 amino acid residues long and have 40% to 50% identities with the N-terminal module of the *S. albus* peptidase, from S14 to Q77. Each has an intervening peptide, 23 amino acid residues long in ORFL3 and 28 amino acid residues long in CwlL, connecting the two sequences described above. The connecting peptides also share about 50% identities with each other (not shown).

(iii) The C-terminal region of the *C. acetobutylicum* lysozyme, from V190 to N324, exhibits, still, another type

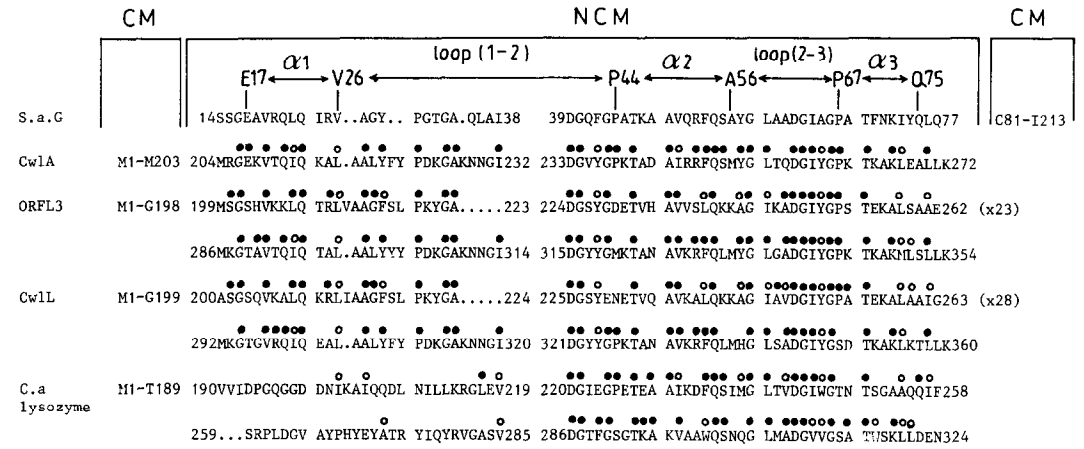

Fig. 5. Amino acid alignments of the non-catalytic, N-terminal module of the *Streptomyces albus* G Zn DD-peptidase (S.a.g.) and the non-catalytic, C-terminal modules of the Bacillus subtilis CwlA amidase, Bacillus licheniformis ORFL3 and CwlL amidases and Clostridium acetobutylicum (C.a.) lysozyme. The secondary structures of the *S. albus* peptidase are indicated. Filled circles identify strict identities by reference to the *S. albus* peptidase. Open circles identified hydrophobic residues occurring at equivalent places. CM = catalytic module. NCM = non-catalytic module.

of design. It consists of two contiguous polypeptide segments V190–F258 and S259–N324. The amino terminal portions, V190–V219 and S259–V285, of each of these segments are variable but the carboxy terminal portions, D220–F258 and D286–N324, each have 50% identities with the carboxy terminal portion, D39–Q77, of the N-terminal module of the *S. albus* peptidase.

(iv) With Pho = hydrophobic amino acid residue and using the amino acid numbering of the *S. albus* peptidase, the general consensus of the eight aligned sequences of Fig. 5 is:

```
39      42          46     50     53  54     59
D G  X  Pho   G   X2  T  X3  Pho X2 Pho  Q X3 G Pho X2
        (E) * (E) *
```

```
62  64  65          69     73     76  77
D G Pho Pho   G   X2  T  X3  Pho X2 Pho  X
                                   (E) *
```

(E)*: at position 229 of repeat 1 in the CwlL amidase; at positions 223 of repeat 1 and 323 of repeat 2 in the *C. acetobutylicum* lysozyme.

Remarkably, D39, D62, T46, T69, Q54 and the hydrophobic amino acid residues at positions 50, 53, 59, 64, 65 and 73 are conserved. As shown in Fig. 4, all these amino acid residues occupy topologically important positions in the cavity of the non-catalytic module of the *S. albus* peptidase.

(v) As a corollary of the alignments shown in Fig. 5, the catalytic modules of the CwlA, ORFL3 and CwlL amidases are about 200 amino acid residues long. They have similarity in their primary structures (not shown) and they lack the GSNRY (or analogue) motif thought to be a general marker of the amidases [18]. The catalytic and non-catalytic modules of the peptidoglycan amidases of known primary structure show variations in hydrophobic cluster patterns, suggesting more than one family of amidases (unpublished data).

3. Discussion

Acquisition of a substrate binding module is an evolutionary advantage for exocellular enzymes that interact with and hydrolyse bonds in a polymeric insoluble substrate. In this respect, the data presented above allow the following conclusions to be drawn. As established by X-ray crystallography studies (and with CM = catalytic module and NCM = non-catalytic module), the modular design of the *S. albus* peptidase is NH_2-NCM-CM-COOH. Its non-catalytic module possesses, exposed at the surface, an elongated cavity that bears a typical amino acid sequence signature in the form of two α-helical repeats, 16 amino acid residues long, connected by a heptapeptide loop. As derived from amino acid alignments, the modular design of the *B. subtilis* CwlA amidase is NH_2-CM-NCM-COOH, that of the *C. acetobutylicum* lysozyme is NH_2-CM-NCM-NCM-COOH and (with IP = an intervening peptide about 25 amino acid residues long) that of the *B. licheniformis* ORFL3 and CwlL amidases are NH_2-CM-NCM-IP-NCM-COOH. Each of these non-catalytic modules possesses a cavity whose shape and structure is comparable to that which is present in the non-catalytic module of the *S. albus* peptidase.

The likely function of these cavities as substrate recognition and binding sites would depend mainly on the glutamine residue, the two aspartic acid residues and the two threonine residues which are strictly conserved and the hydrophobic residues which occur at equivalent places along the cavity-defining motif. The peptidase, lysozyme and amidases under study hydrolyse different bonds in the bacterial cell wall peptidoglycan. Variations in the sequences of the binding motif due to the occurrence of non-conserved amino acid residues could be responsible for substrate specificities and topologically directed activities. Full characterization of the non-cata-

lytic modules as binding sites, however, requires identification of the recognized polymer and study of the ligand-binding site interactions by genetic engineering, molecular modelling and site-directed mutagenesis.

Acknowledgements: This work was supported in part by the Belgian programme on Interuniversity Poles of attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming (PAI No. 19) and the Fonds de la Recherche Scientifique Médicale (Contract No. 3.4531.92).

References

[1] Shockman, G.D. and Höltje, J.V. (1994) in: Bacterial Cell Wall (Ghuysen, J.M. and Hakenbeck, R. Eds.) New Comprehensive Biochemistry 27, pp. 131–136, Elsevier, Amsterdam, New York.
[2] Kehoe, M.A. (1994) in: Bacterial Cell Wall (Ghuysen, J.M. and Hakenbeck, R. Eds.) New Comprehensive Biochemistry 27, pp. 217–261, Elsevier, Amsterdam, New York.
[3] Lopez, R., Garcia, J.L., Garcia, E., Ronda, C. and Garcia, P. (1992) FEMS Microbiol. Lett. 100, 439–448.
[4] Diaz, E., Lopez, R. and Garcia, J.L. (1991) J. Biol. Chem. 266, 5464–5471.
[5] Garcia, P., Garcia, J.L., Garcia, E. and Lopez, R. (1986) Gene 43, 265–272.
[6] Garcia, P., Garcia, J.L., Garcia, E., Sanchez-Puelles, M and Lopez, R. (1990) Gene 86, 81-88.
[7] Duez, C., Lakaye, B., Houba, S., Dusart, J. and Ghuysen, J.M. (1990) FEMS Microbiol. Lett. 71, 215–220.
[8] Croux, C. and Garcia, J.L. (1991) Gene 104, 25–31.
[9] Kuroda, A. and Sekiguchi, J.J. (1990) J. Gen. Microbiol. 136, 2209–2216.
[10] Foster, J.J. (1991) J. Gen. Microbiol. 137, 1987–1998.
[11] Oda, Y., Nakayama, R., Kuroda, A. and Sekiguchi, J. (1993) Mol. Gen. Genet. 241, 380–388.
[12] Lee, J.W.K., Edwards, C.W. and Hulett, F.M. (1991) J. Gen. Microbiol. 137, 667–677.
[13] Ghuysen, J.M., Dierickx, L., Coyette, J., Leyh-Bouille, M., Guinand, M. and Campbell, J.N. (1969) Biochemistry 8, 213–222.
[14] Ghuysen, J.M., Leyh-Bouille, M., Bonaly, R., Nieto, M., Perkins, H.R., Schleifer, K.H. and Kandler, O. (1970) Biochemistry 9, 2955–2961.
[15] Dideberg, O., Charlier, P., Dive, G., Joris, B., Frère, J.M. and Ghuysen, J.M. (1982) Nature 299, 469–470.
[16] Wéry, J.P. (1987) Ph.D. Thesis, University of Liège, Belgium.
[17] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[18] Lazarevic, V., Margot, P., Soldo, B. and Karamata, D. (1992) J. Gen. Microbiol. 138, 1949–1961.