Sequence-imposed structural constraints in the TonB protein of *E. coli*

J.S. Evans†, B.A. Levine*, I.P. Trayer+, C.J. Dorman° and C.F. Higgins°

Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, +Department of Biochemistry, University of Birmingham, PO Box 363, Birmingham B15 2TT, England and *°Molecular Genetics Laboratory, Department of Biochemistry, University of Dundee, Dundee DD1 4HN, Scotland*

Received 4 September 1986

The solution conformation of a 33-residue peptide segment, derived from the TonB protein which is implicated in bacterial membrane transport processes, has been investigated using high-resolution proton magnetic resonance techniques. This proline-rich peptide possesses sequence-imposed sections of elongated secondary structure that must be retained in the native protein configuration. These structural constraints provide elements of stiffness that imply a purely structural role for TonB and are relevant to the subcellular location and biological role of the protein. On the basis of these data we suggest that this protein spans the periplasmic space, linking the inner and outer membrane components of TonB-dependent transport systems.

TonB protein *Membrane protein* *Protein structure* *Proline* *^{1}H-NMR*

1. INTRODUCTION

The *tonB* gene in *E. coli* encodes an unusually proline-rich protein of 242 amino acids. Although the precise function of this protein remains obscure, it is known to be required for several specific membrane transport processes, including the uptake of vitamin B12 and iron-chelator complexes as well as for productive infection by a number of bacteriophages [1]. Despite its unusual amino acid composition and hydrophilic character, the TonB protein is membrane-associated [1,2]. We have recently determined the *tonB* gene sequence of *S. typhimurium* [2], and the predicted protein sequence shows a very high degree of homology with that of *E. coli* [1].

* To whom correspondence should be addressed

† On sabbatical leave from Lawrence University, Appleton, WI 54912, USA

The unusually high proline content of the TonB proteins from these two species (41 of 242 residues, or 17%) is particularly reflected in a 33-residue sequence which is totally conserved between the two species:

1 9 10 11 12 13 14 15 16
$(Glu\text{-}Pro)_4$-Ile-Pro-Glu-Pro-Pro-Lys-Glu-Ala-

17 18 19 20 21 33
Pro-Val-Val-Ile-Glu-$(Lys\text{-}Pro)_6$

This 33-residue segment occurs some 70–75 amino acids from the N-terminus of the protein as determined from the gene and contains repeating -Glu-Pro- and -Lys-Pro- dipeptide units connected by a section composed of a very limited range of residue types: Glu, Lys, Ala, Ile, Val, and Pro. We have used ^{1}H NMR techniques to study the structural features associated with this unusual 33-residue segment of the TonB protein.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

2. EXPERIMENTAL

The 33-residue polypeptide was prepared using a programmed synthesizer and was purified by a combination of gel filtration on Sephadex G-25 and ion-exchange chromatography on SP-Trisacryl. Purity was checked by amino acid analysis and high-performance liquid chromatography. The N-terminus of the synthesized molecule has a free amino group and the C-terminus is amidated.

^1H NMR experiments were conducted using facilities of the Oxford Enzyme Group, including Bruker WH-300 and AM-500 spectrometers as well as a 500-MHz instrument with Nicolet computer and software; all of the NMR spectrometers have cryomagnets from Oxford Instruments. Data sets from two-dimensional homonuclear correlated spectroscopy (COSY) and nuclear Overhauser effect spectroscopy (NOESY) were analysed using contour plots and one-dimensional plots of subspectra.

Proton chemical shift values were determined relative to sodium 3-(trimethylsilyl)-1-propanesulfonate (TSS) as internal reference and generally agreed within ± 0.02 ppm between spectra recorded under similar conditions at 300 or 500 MHz. The solutions studied contained no added buffers. Solutions in ^1H_2O contained added ^2H_2O to provide a spectrometer lock signal. All pH measurements were direct meter readings.

3. RESULTS

Evidence of local secondary structure for the backbone of the 33-residue TonB peptide derives directly from the observation of differential resonance positions for backbone $C^{\alpha}H$ protons for each of several residue types (fig.1), as well as from the spread in chemical shift of exchangeable backbone NH proton signals (see below). The various backbone and side chain signals were assigned to amino acid type using homonuclear correlated spectroscopy. In contrast to protons attached to the backbone, the majority of protons in side chains resonate at chemical shift positions that may be associated with a largely solvent environment with the marked exception of the prolyl $C^{\delta}H_2$

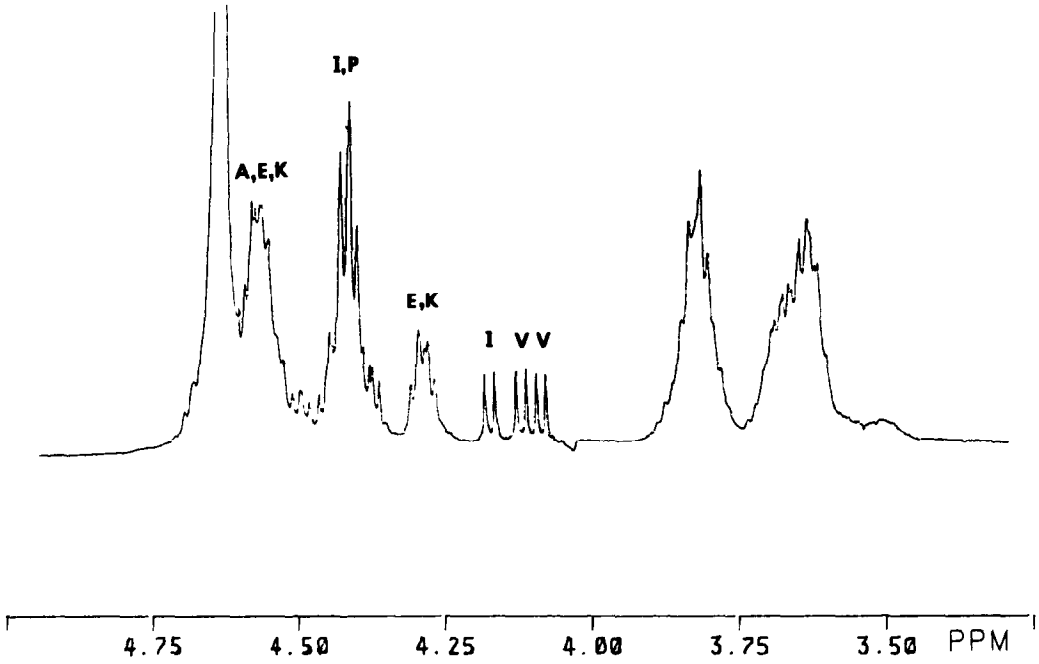

Fig.1. ^1H NMR spectrum of TonB peptide in ^2H_2O (pH 7.7, 310 K) showing $C^{\alpha}H$ (4.6–4.0 ppm) and prolyl $C^{\delta}H_2$ (3.8, 3.6 ppm) resonances. Assignments to particular residue types (one-letter abbreviations for amino acids) are based on analysis of spin coupling patterns observed in decoupling and COSY experiments.

protons. The latter give rise to two resonance envelopes, roughly equal in intensity, at 3.82 and 3.65 ppm (fig.1).

The two prolyl $C^{\delta}H_2$ signals are strongly coupled to one another (fig.2A), showing that each signal must be ascribed to superposition of resonances from one of the inequivalent protons at the C^{δ} position within most if not all of the proline residues, rather than to either *cis-trans* isomerism or some hypothetical dichotomy of proline residues in the sequence. Previous spectral studies of similar proline-containing segments derived from the pyruvate dehydrogenase complex of *E. coli* [3] and the N-terminal region of the LC1 alkali light chain of skeletal myosin [4] have led to the correlation of such spectral inequivalence with all-*trans* X-Pro linkages. The latter study of the -Ala-Pro- repeats of the alkali light chain identified the corresponding alanyl $C^{\alpha}H$ signal at 4.56 ppm, downfield of the other alanyl resonances, while strong nuclear Overhauser effects (NOE) among the tightly packed alanyl $C^{\alpha}H$ and prolyl $C^{\delta}H_2$ protons indicated inter-proton distances of approx. 0.2 nm.

Similarly, two-dimensional NOESY experiments with the TonB peptide show a strong NOE between the prolyl $C^{\delta}H_2$ protons and the $C^{\alpha}H$ protons which resonate at 4.57 ppm (fig.2B). Therefore we can distinguish the X-Pro and X-Y $C^{\alpha}H$ signals for glutamyl and lysyl residues (fig.1) as being those at 4.6 ppm (X-Pro) and 4.3 ppm (X-Y), respectively, in agreement with the coupling patterns obtained

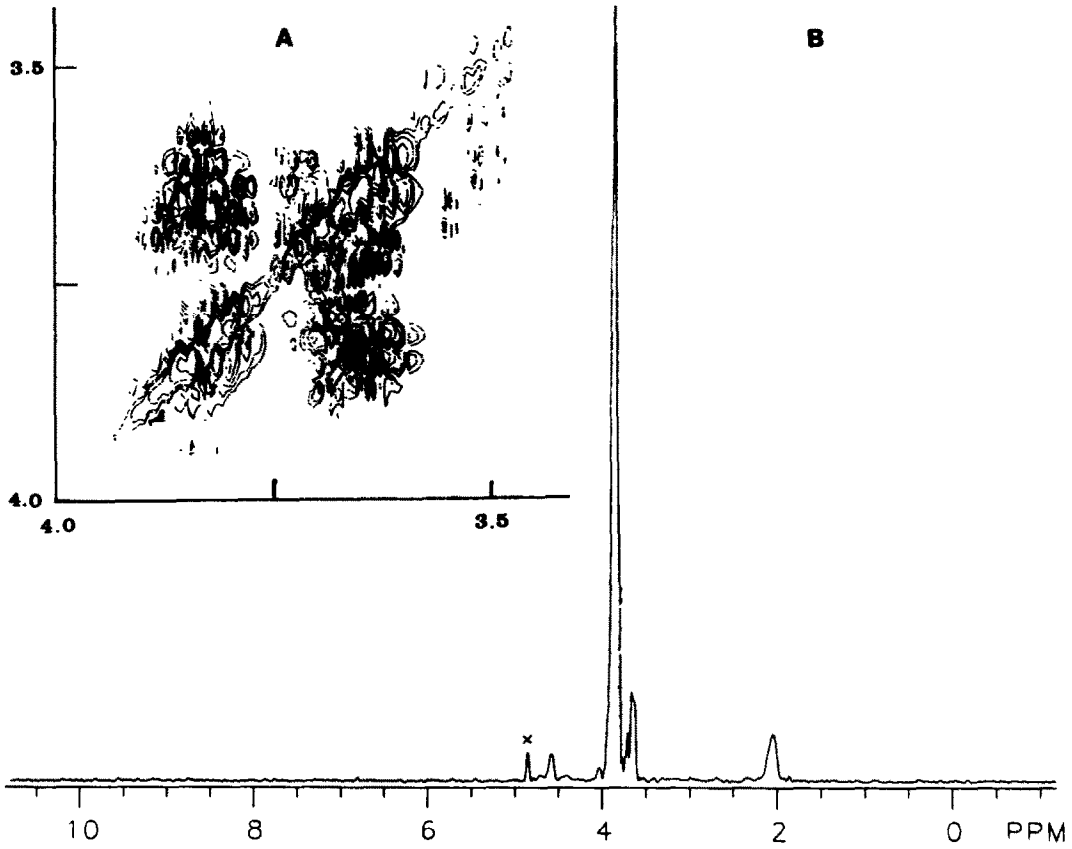

Fig.2. ^1H NMR properties of the $C^{\delta}H_2$ protons of the proline residues. (A) Contour representation of 500-MHz phase-sensitive COSY spectrum of TonB peptide in ^2H_2O (pH 7.7, 310 K) in the region showing spin coupling between inequivalent $C^{\delta}H_2$ protons of the proline residues. (B) Sub-spectrum from 500-MHz two-dimensional NOESY data set showing NOE from the resonance position of one $C^{\delta}H_2$ proton (3.8 ppm) to the resonance positions of the other $C^{\delta}H_2$ proton (3.65 ppm), the prolyl $C^{\gamma}H_2$ protons (2.05 ppm), and $C^{\alpha}H$ protons of preceding residues (4.57 ppm).

from analysis of the homonuclear COSY data. This observation of the strong NOE between each of the prolyl $C^{\delta}H_2$ protons and the backbone $C^{\alpha}H$ protons of the preceding residues also demonstrates the presence of a predominantly *trans* configuration for X-Pro linkages in the repeating -Glu-Pro- and -Lys-Pro- dipeptide units in the sequence as well as for the linkages involving Ile-9, Pro-12, and Ala-16. One-dimensional NOE experiments carried out at 300 MHz have shown these various $C^{\alpha}H$-$C^{\delta}H_2$ NOE to be of negative sign at the lower frequency as at 500 MHz and thereby confirm the closely packed, elongated (*trans*) backbone configuration that extends over a large part of the proline-rich primary structure.

Further evidence of sequence-induced constraints on the configuration of this segment of the TonB protein is provided by the backbone NH resonances observed in ^1H_2O in COSY, NOESY, and one-dimensional experiments. Patterns of spin coupling to side-chain protons and one-dimensional decoupling experiments have established assignments of $C^{\alpha}H$ and NH resonances. We concentrate here on the data for alanyl, valyl, and isoleucyl residues (fig.3A). The corresponding NOESY contour plot (fig.3B), indicative of inter-proton proximity, shows a conspicuous column of strong inter-residue cross-peaks at 4.4 ppm between prolyl $C^{\alpha}H$ protons and NH protons of succeeding residues, i.e. Pro-Y. One such inter-residue cross-peak connects $C^{\alpha}H$ of a proline to NH of a valine and thus identifies Val-18, whose much less intense intra-residue NOE can also be discerned. In turn, a cross-peak connects the $C^{\alpha}H$ resonance of Val-18 to the NH resonance of Val-19, another from Val-19 to Ile-20, and so forth for several successive linkages; this succession of close inter-residue proton-proton

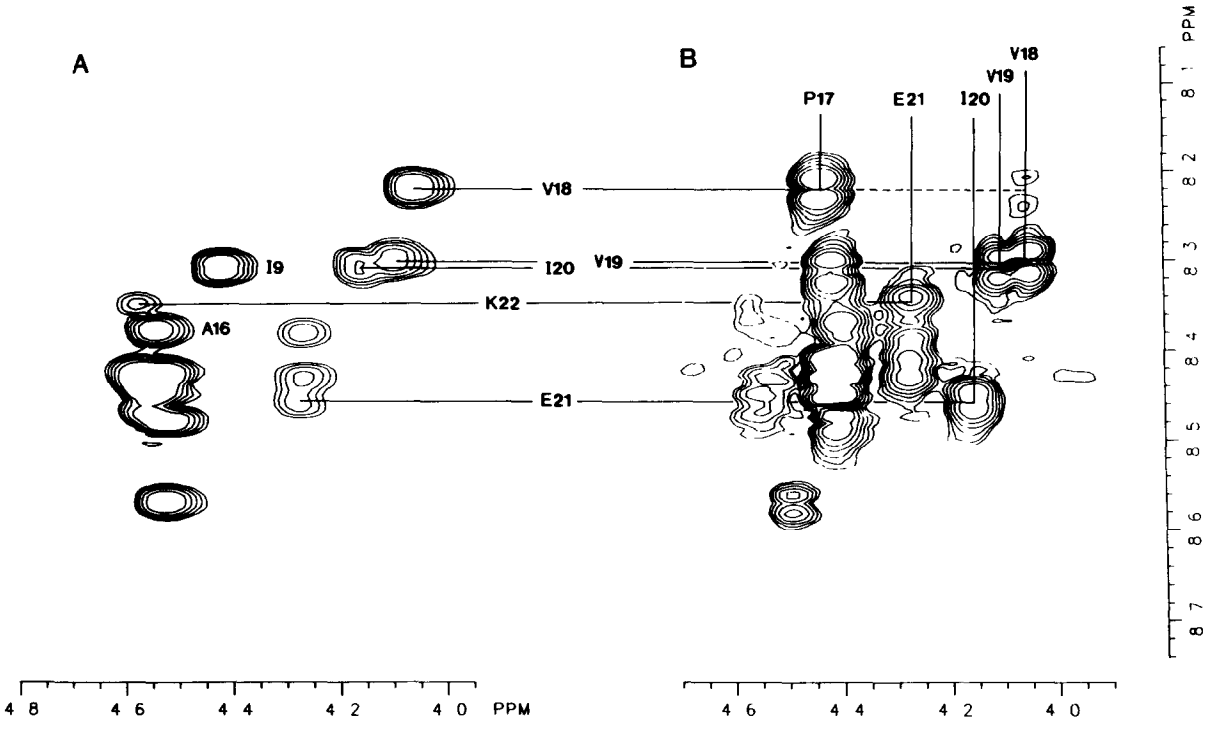

Fig.3. Contour representation of 500-MHz two-dimensional ^1H NMR spectra of TonB peptide in ^1H_2O (pH 7.7, 290 K). (A) COSY spectrum, in which cross-peaks are expected for the spin-coupled $C^{\alpha}H$ and NH signals from all residue types except proline; (B) NOESY spectrum (mixing time 120 ms with randomization of 10 ms), in which a strong cross-peak signifies that certain $C^{\alpha}H$ and NH protons (either intra- or inter-residue) are closer than ~0.25 nm. Horizontal lines correlate certain NH resonance positions from the COSY spectrum at the $C^{\alpha}H$ resonance positions of the respective preceding residues in the primary sequence, as identified by vertical lines. The COSY cross-peaks for some other residues are labelled.

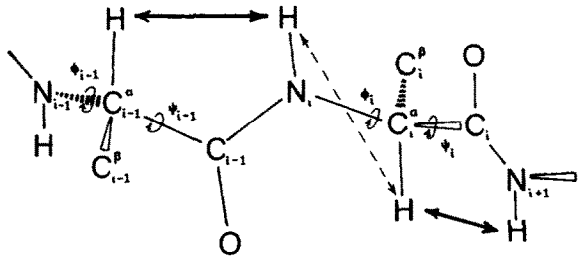

Fig.4. Diagram of peptide backbone geometry which is consistent with the observation of much stronger inter-residue NOE from $C^{\alpha}H_{i-1}$ to NH_i than intra-residue NOE from $C^{\alpha}H_i$ to NH_i.

contacts appears to continue as far as $C^{\alpha}H$ of Glu-21 to NH of Lys-22.

4. DISCUSSION

The 33-residue peptide from the TonB protein contains three distinguishable structural elements, two rod-like sections of different charge whose shape is determined by multiple *trans* Glu-Pro or Lys-Pro repeats, separated by a six-residue section resembling a β-strand. This middle section of the TonB peptide, Pro^{17}-Val-Val-Ile-Glu-Lys^{22}, which contains successive residues with bulky, branched, or charged side chains, shows strong inter-residue backbone $C^{\alpha}H$-NH NOE, indicating that it must adopt an extended backbone structure that is well-defined in conformational space. Each NH proton in this section of the peptide is disposed much closer to the $C^{\alpha}H$ proton of the preceding residue than to that of its own residue, thereby placing limits to the excursions possible for the backbone torsion angles. The idealised local geometry which meets this requirement (fig.4), bringing the inter-residue distance between $C^{\alpha}H$ and NH protons near the minimum (0.22 nm) and the intra-residue distance near the maximum (0.28 nm), resembles the rather elongated backbone in β-strands in proteins [5]. The branched side chains of valine and isoleucine tend to constrict the potential range of torsion angles [6] and are indeed frequently observed to occur in β-sheet rather than helical regions in proteins. In this respect it is of interest that the proline-rich TonB protein of *E. coli* contains another similar cluster of bulky and branched side chains, namely the residue sequence Ile-Val-Val-Asn-Ile-Leu near its C-terminus [1].

The configuration resembling a β-strand adopted by residues 17–22 of the TonB peptide is flanked by similarly extended proline-rich backbone sections. These flanking sections, composed of repeated *trans* X-Pro dipeptide units, adopt relatively stiff elongated conformations because of packing constraints imposed upon backbone torsion angles. Closure of the pyrrolidine ring determines ϕ for proline, and steric repulsions between its $C^{\delta}H_2$ group and the C^{β} substituents of X restrict ψ for the preceding X residues [7], as demonstrated here and in [4]. Moreover, although each of the values $\psi \approx -50°$ and $\psi \approx +130°$ is feasible for an isolated *trans* proline residue in a peptide chain [7], the former value is excluded on the basis of the data presented here because of the close proximity between the $C^{\alpha}H$ of each X and the $C^{\delta}H_2$ of the succeeding proline in *trans* X-Pro units, whose backbone fluctuations ($\pm$ 30°) are highly correlated. Further, the observation of NOE between the backbone NH of X and the $C^{\alpha}H$ of the preceding proline (inter-proton distance $\leq$0.25 nm) imposes limits on excursions of the ϕ torsion angle for the X residue (e.g. Glu or Lys) of each Pro-X pair in the repeated sections. Thus for each X-Pro residue pair within those sections, instead of the usual situation of some considerable freedom of motion involving two ϕ and two ψ torsion angles per residue pair, only the angle ϕ for X remains free from sequence-imposed constraints on its potential range. Based on a C^{α}–C^{α} distance of 0.38 nm, each of these flanking sections of X-Pro repeats is some 4–5 nm in length, which is sufficient to span the periplasmic space [8].

The stereochemical basis for the observed structural preference is intrinsic to the amino acid sequence, and these structural features must therefore be retained in the native TonB protein molecule. Indeed, the elongated overall conformation shown by NMR for the 33-residue peptide derived from the N-terminal third of the TonB protein is entirely consistent with the anomalous migration of the molecule on SDS-polyacrylamide denaturing gels [1,9]. The remarkable residue composition (~30% proline) and sequence of the N-terminal half are observed to impose well-defined structural constraints and confer a shape and

overall dimension (up to 10 nm, see above) which suggest that the primary role of the TonB protein is structural, spanning the periplasm and linking the inner and outer membranes. Except for the putative signal peptide sequence at the N-terminus, only one short region of the primary structure is notably hydrophobic in character. This latter region, at the C-terminal end of the molecule, is presumably involved in anchoring the protein to the inner membrane. The otherwise hydrophilic protein is therefore likely to be exposed on one side or other of the membrane and, using antibodies raised against the 33-residue peptide, we have shown that at least this part of the protein is periplasmic [2].

The importance of the intrinsic chain stiffness imposed by the *trans* X-Pro linkages is here to restrict the segmental flexibility of the backbone to enable specific, large-scale changes in spatial orientations. The section connecting the two elongated X-Pro structures contains specific linkages with much greater segmental freedom, Lys-Glu-Ala, where swivelling motions may occur. Together with the conformational adaptability of a β-strand, these various structural features suggest that the involvement of the TonB protein in transport processes is related to its conformationally endowed ability to link distinct receptors whose functions are localised in space. This view is entirely consistent with the biological roles of TonB [1,2]. Most substrates are sufficiently small to pass through the outer membrane porins and hence do not require a specific outer membrane receptor. Transport systems for such substrates are *tonB*-independent. TonB is only required for those few transport processes (B12, iron chelates, phage DNA) which involve both outer and inner membrane receptors. The elongated shape of the TonB protein has presumably evolved to facilitate the coordination of spatially distinct sites and thereby maximise free energy coupling. Similar correlation of such sequence-imposed elongated structures and the resulting efficiency in lateral diffusion has been described for the pyruvate dehydrogenase complex [10] and the alkali light chain of myosin [4].

ACKNOWLEDGEMENTS

We thank Professor R.J.P. Williams and Dr G.C. Barr for helpful discussions and Sue Brewer for providing the purified polypeptide. This work was supported by grants from the SERC and MRC. C.F.H. is a Lister Institute Research Fellow.

REFERENCES

[1] Postle, K. and Good, R.F. (1983) Proc. Natl. Acad. Sci. USA 80, 5235–5239.
[2] Dorman, C.J., Barr, G.C., Lecocq, J., Davilson, M.T., Trayer, I.P. and Higgins, C.F. (1986) EMBO J., in preparation.
[3] Radford, S.E., Laue, E.D. and Perham, R.N. (1986) Abstracts of Cambridge Meeting, Biochemical Society, p.152.
[4] Bhandari, D.G., Levine, B.A., Trayer, I.P. and Yeadon, M.E. (1986) Eur. J. Biochem., in press.
[5] Wuthrich, K., Billeter, M. and Braun, W. (1984) J. Mol. Biol. 180, 715–740.
[6] Nemethy, G., Leach, S.J. and Scheraga, H.A. (1966) J. Phys. Chem. 70, 998–1004.
[7] Flory, P.J. (1969) Statistical Mechanics of Chain Molecules, pp.268–274, Wiley, New York.
[8] Hobot, J.A., Carleman, E., Villiger, W. and Kellerberger, E. (1984) J. Bacteriol. 160, 143–152.
[9] Postle, K. and Reznikoff, W.S. (1979) J. Mol. Biol. 131, 619–636.
[10] Packman, L.C., Hale, G. and Perham, R.N. (1984) EMBO J. 3, 1315–1319.