Volume 258, number 1, 127–132 FEB 07844 November 1989

Sequence analysis of the cloned streptococcal cell surface antigen I/II

C. Kelly, P. Evans, L. Bergmeier, S.F. Lee*, A. Progulske-Fox*, A.C. Harris+, A. Aitken+, A.S. Bleiweis* and T. Lehner

*Department of Immunology, United Medical and Dental Schools of Guy's and St Thomas's Hospitals, London Bridge, London SE1 9RT, England, *Department of Oral Biology, University of Florida, Gainesville, FL 32610, USA* and *+Laboratory of Protein Structure, National Institute for Medical Research, The Ridgeway, London NW7 1AA, England*

Received 25 September 1989

The gene *spa P* (formerly designated as *spa* P1) encoding the M_r 185,000 surface antigen (I/II) of *Streptococcus mutans*, serotype c (NG5), has been sequenced. The gene (4683 bp) encodes a protein of 1561 amino acid residues including putative signal peptide (residues 1–38) and transmembrane (residues 1537–1556) sequences. The N-terminal region (60–550) has alanine-rich repeats and is predicted to be α-helical. However, the C-terminal region (800–1540) is proline-rich and favours an extended structure. Except for a short central variable region the sequences appear to be highly conserved for *S. mutans* serotype c. N-Terminal sequencing of separated antigen I and antigen II polypeptides suggests that the former represents the N-terminal and the latter the C-terminal portions of the intact antigen.

Antigen I/II; Streptococcal surface antigen; *spa* P1

1. INTRODUCTION

Streptococcus mutans is a microorganism commonly found in the human oral cavity and is responsible for dental caries [1]. Investigations of the protein composition of *S. mutans* revealed 4 antigens: I, II, III and I/II [2]. Of these, antigen I/II [3] is of particular significance, as it is a cell surface [4,5], immunodominant antigen which elicits serum IgG antibodies [6] and T cell helper or suppressor functions [7]. The antigen is a glycoprotein of approximate $M_r = 185,000$, also termed antigen B [8] and P1 [9] and may function as a virulence factor [10]. Immunisation studies with the whole organism [11,12], as well as the M_r 185,000 antigen I/II prevented colonisation of *S. mutans* and the development of dental caries in non-human primates [6,13.] It was suggested that antigen I/II may function as an adhesin by virtue of its hydrophobicity [14]. Indeed, construction of antigen I/II deficient mutants by insertional inactivation to the *spa P* gene has established that the cell surface hydrophobicity of the mutant organism is decreased [15] and that this surface protein functions as an adhesin for binding to salivary agglutinin coated hydroxyapatite beads [16]. Furthermore, monoclonal antibodies raised against the streptococcal antigen I/II [17] were applied in local passive immunisation experiments in non-human primates and prevented colonisation of *S. mutans* and the development of dental caries [18]. This is consistent with the concept that antigen I/II functions as an adhesin. An extensive series of in vivo studies in human subjects has established that some monoclonal antibodies to antigen I/II can prevent colonisation of *S. mutans* and that this is a function of the $F(ab)_2$ fragment of the antibody [19,20].

Antigen I and antigen II, although both present in the intact surface antigen of $M_r = 185,000$, are serologically distinguishable and have been separated biochemically [3,21]. Antigen I was identified as a polypeptide of approximate $M_r = 150,000$ which was purified from culture supernatants of *S. mutans*. In contrast, antigen II was only separated following proteolysis of antigen I/II (a process which destroyed antigen I) as a polypeptide of approximate $M_r = 50,000$. Recently, cloning of the gene, termed *spa P*, encoding antigen I/II (P1) of *S. mutans* strain NG5 (serotype c) was reported [22]. The cloned *spa P* was present in a 5.2 kb fragment derived by partial digestion of genomic DNA with Hind III. Cloning of the identical gene from another *S. mutans* strain (MT8148) was also reported [23].

In this report the complete nucleotide sequence of *spa P* is presented, together with the predicted amino acid sequence. N-Terminal amino acid sequencing of the purified antigens I and II suggests that antigen I is present in the N-terminal and antigen II in the C-terminal portions of the M_r 185,000 antigen.

Correspondence address: C. Kelly, Department of Immunology, United Medical and Dental Schools of Guy's and St Thomas's Hospitals, London Bridge, London SE1 9RT, England

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

2. MATERIALS AND METHODS

2.1. *Nucleotide sequencing*

Restriction fragments from the recombinant clone pSM2949 [22] were subcloned into M13 mp18 or M13 mp19. The sequence of *spa P* was determined on both strands using the dideoxy chain termination procedure [24]. Double stranded PCR products (see below) were sequenced directly using a modified procedure [25]. Synthetic primers (17–20 nucleotides) were used to sequence fragments > 400 bases. Sequences were assembled and analysed using the Staden plus programme (Amersham, England).

2.2. *Polymerase chain reaction (PCR)*

Chromosomal DNA was prepared from *S. mutans* NG5 as described previously [22]. Using the primers CAAATGGGACAAACAGGC (nucleotides 3985–4002 from the NG5 sequence) and AAGGCAGTGCGAAGTACC (complementary to nucleotides 5153–5170 from the MT8148 sequence [23]) a 1 kb fragment which included the 3′ region of *spa P* was amplified by the PCR [26] (30 cycles: 94°C for 2 min, 55°C for 1 min, 72°C for 2 min).

2.3. *Protein purification and sequence analysis*

Antigen I/II and the separated antigens I and II were purified from *S. mutans* serotype c (Guy's strain) as described [3,21]. A peptide of $M_r = 4000$ was isolated from a subtilisin digest of antigen I/II [27]. A recombinant antigen of $M_r = 155{,}000$ was prepared by gel filtration of the periplasmic fraction, released following osmotic shock from *E. coli* LC 137 transformed with pSM2949 [22]. Purified proteins (100–200 pmol) were subjected to Edman degradation, using the 470A gas phase sequencer with on line 120A PTH analyser (Applied Biosystems).

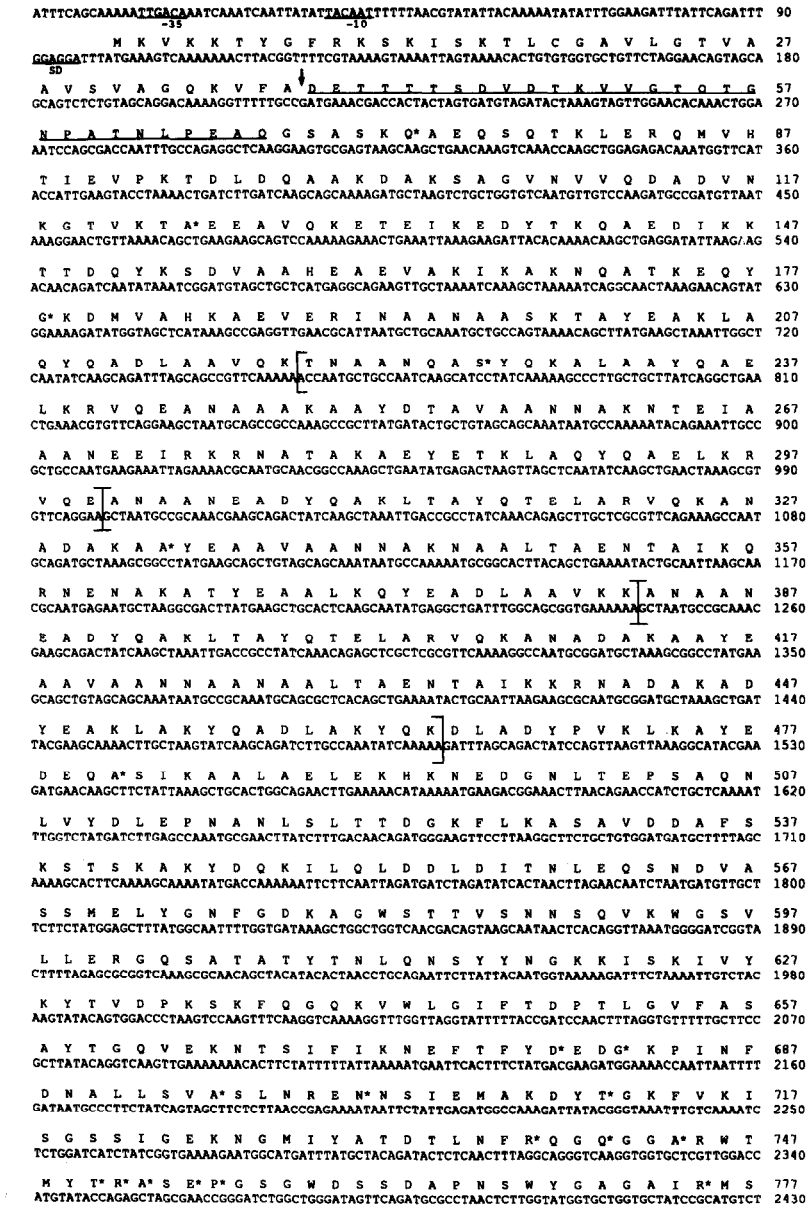
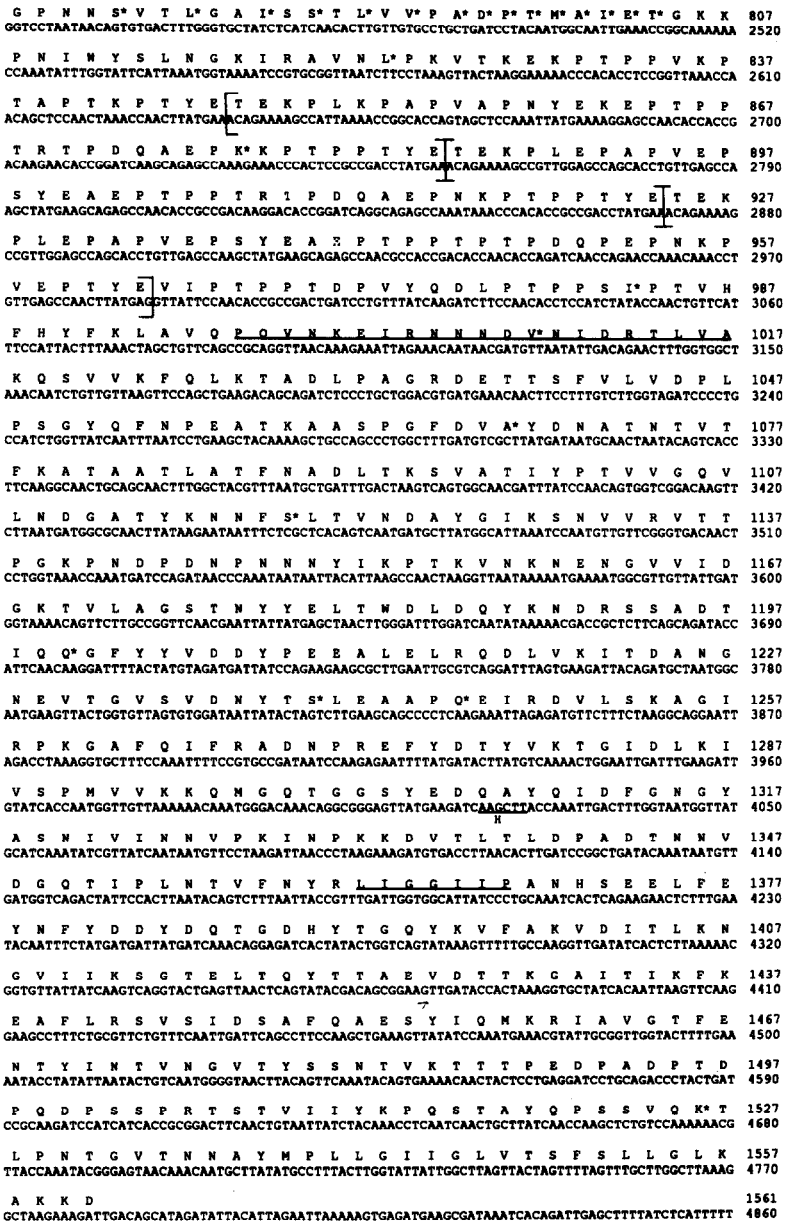


Fig.1. Sequence of *spa P* from *S. mutans*. The most probable −35 and −10 promoter regions and Shine-Dalgarno sequence (SD) are underlined as is the 3 *Hind* III site (H) of pSM2949. Amino acid residues underlined are those determined by direct protein sequencing. Repeating amino acid

3. RESULTS

The complete nucleotide sequence of the gene (4683 bp) encoding antigen I/II and the deduced amino acid sequence is shown in fig.1. An open reading frame from position 100–4782, as shown, was identified coding for a polypeptide of 1561 residues. The most probable −10 and −35 promoter sequences are underlined as is a putative Shine-Dalgarno sequence. The full sequence of the pSM2949 insert includes approximately 800 bases 5′ to that shown in fig.1. In order to confirm the reading frame, N-terminal amino acid sequence analyses were performed on intact native antigen I/II, recombinant gene products, as well as a number of fragments derived from antigen I/II with the results shown in fig.2. In each case the amino acid sequence determined matched that predicted from the nucleotide sequence. The N-terminal sequence of the recombinant M_r 155,000 antigen I/II corresponded to the predicted sequence from Asp-39 to Gln-67. Similarly the N-terminal residues of a purified M_r 185,000 antigen I/II and a M_r 150,000 antigen I showed the same partial sequence, suggesting that the N-terminal 38 residues represent a signal peptide. The predicted polypeptide M_r of the mature protein, after the signal peptide cleavage, is 166,159 which is lower than the value determined by SDS-PAGE. The



sequences are bracketed. Amino acid substitutions relative to the sequence of *S. mutans* strain MT8148 are marked(*). A potential signal peptide cleavage site is arrowed. The N-terminal residue of antigen I is Asp-39 and of antigen II is Pro-997 (also indicated in fig.2).

	1 10 20 29	POSITION IN SEQUENCE
Antigen I/II(185,000)	- E T T T	39-43
Antigen I (150,000)	D E T T - - S	39-45
Rec 155,000 (E.coli)	D E T T T T S D V D T K V V G T Q - G N P A T N L P E A Q	39-67
Antigen II (50,000)	- P Q V N - E I - N N N - V N I D - T - - A	997-1017
4,000 M, Peptide (from subtilisin digestion)	- L I G G I - P	1361-1368

Fig.2. N-Terminal sequence of antigen I/II and derived peptides. Recombinant antigen I/II (rec 155,000) was purified from *E. coli* transformed with plasmid pSM2949. Residues are numbered as for the sequence of strain NG5. - indicates no residue identified.

difference may be due to glycosylation or to the presence of the Pro-rich sequences which would result in anomalous electrophoretic migration.

Comparison with the sequence of Okahashi et al. [23] suggested that the complete reading frame comprised a further 800 bases beyond the 3′ *Hind* III site. By means of the PCR, and using primers based on the sequences of *S. mutans* strains NG5 and MT8148, a 1 kb DNA fragment which overlapped with the pSM2949 insert was generated from *S. mutans* NG5 DNA. Sequencing of the PCR product confirmed that it encoded the entire C-terminal region of antigen I/II, as evidenced by the presence of a termination codon at nucleotides 4783–4785.

The predicted amino acid sequence of antigen I/II includes a potential transmembrane region located at the C-terminus of the protein (Ala-1537–Leu-1556). Immediately preceding this is a sequence rich in Pro and otherwise consisting almost entirely of polar residues (Thr-1486–Asn-1536). Similar sequences present in other streptococcal cell surface proteins are believed to span the cell wall [28]. Residues 847–963 represent 3 tandem repeats of a Pro-rich 39 residue sequence and residues 219–464 represent 3 tandem repeats of an Ala-rich sequence. Residues 136–218 may represent a further degenerate repeat of the latter sequence. These results are consistent with those for the *S. mutans* strain MT8148 gene sequence [23].

The sequences from the two *S. mutans* strains (NG5 and MT8148), are highly conserved, with most nucleotide substitutions reflecting the use of alternative codons or resulting in conservative amino acid substitutions. There are 36 single amino acid substitutions distributed throughout the protein. Two short stretches within the region encoding amino acid residues 750–805, however, showed considerably more variation. Sequence comparison between these strains and with the corresponding region of the analogous surface antigen from *S. sobrinus* (*spa A*, PAg) [29], as shown in fig.3, indicates that in the region 750–756 a single residue deletion has occurred in strain MT8148, as well as 4 substitutions. In the region 795–805 a sequence of

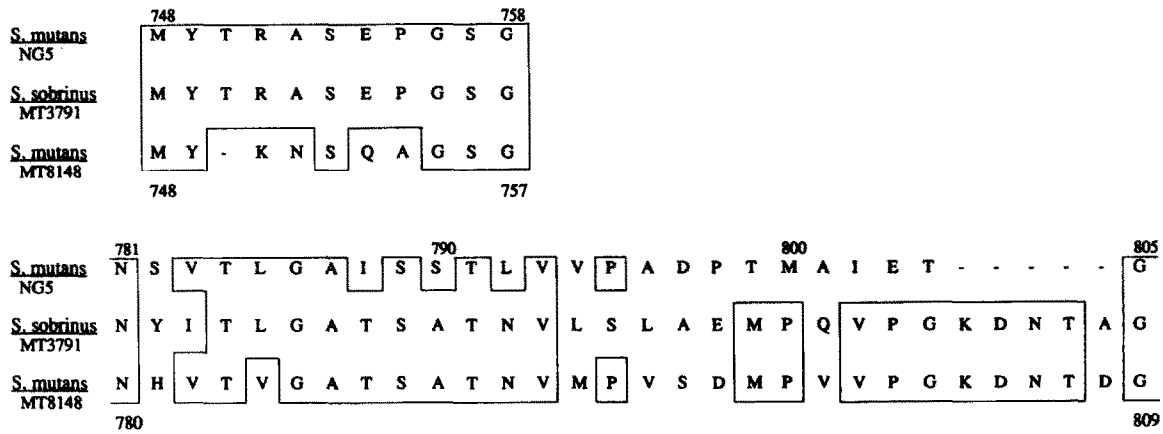


Fig.3. Sequence comparison of variable region of antigen I/II from *S. mutans* strains NG5 and MT8148 and antigen PAg (*spa* A) from *S. sobrinus* strain MT3791 [28]. Identical amino acid residues are boxed.

9 amino acid residues in strain NG5 is replaced with a non-homologous sequence of 14 residues in strain MT8148.

From the N-terminal sequences, the positions of antigen I and antigen II were identified. Thus, antigen I extends from the N-terminus of the M_r 185,000 surface antigen. Depending on the contribution of any post-translational modifications to the M_r it may extend to residues 1390–1400 (see below), and would therefore show some overlap with the antigen II region. Antigen II extends from residue 997 and thus forms most or all of the C-terminal region.

4. DISCUSSION

The predicted sequence of antigen I/II is consistent with that of a cell surface protein, anchored in the membrane by the C-terminal region. The particularly high level of release of antigen I/II into culture supernatants, observed for the NG5 strain [30], is not therefore due to the presence of a truncated gene. Release of antigen I into culture supernatants of *S. mutans* is presumably the result of a cleavage at the C-terminal region by endogenous proteases. Similarly release of an N-terminal fragment was reported for the wall-associated antigen III (protein A) [31], although in neither case is the functional significance, if any, clear. The finding that the subtilisin cleaved M_r 4000 peptide reacts with rabbit antisera against antigen I [27] suggests that this peptide is derived from the C-terminal region of antigen I. Thus, antigen I may well extend as far as residues 1390–1400 and potential cleavage sites would be the clusters of Lys located between residues 1396 and 1440.

The N-terminal region of the protein which includes the Ala-rich repeats between residues 60 and 550 is predicted to be α-helical [32]. The relative abundance of Pro in the C-terminal region, i.e. residues 800–1540, would preclude the formation of a tightly folded structure but would favour a rather extended structure. The sequence which shows variation between the two strains (750–805) falls between these two regions and may represent a conformationally less constrained part of the molecule. Takahashi et al. [29] have shown significant homology between PAg from *S. sobrinus* and the analogous antigen (PAc, I/II) in *S. mutans* to be restricted to residues 612–919 of the *S. mutans* molecule. This region should include cross-reactive epitopes accounting for the reported immunological relationships between these organisms.

In view of the interest in antigen I/II as a component of an anti-caries vaccine, the demonstration that there is limited strain variation is of some importance. In this respect antigen I/II shows similar properties to the wall-associated protein antigen III (A) [31]. Cross-reactivity between *S. mutans* and heart tissue has been raised as a serious objection to vaccination [33]. Later publications claimed that the heart cross-reactive antigen resides in the M_r 185,000 streptococcal antigen I/II [9,13,34]. However, immunisation with antigen I/II of non-human primates failed to elicit heart cross-reactive antibodies [35]. Indeed, little homology was found between the sequence of antigen I/II and that of M protein in *S. pyogenes* causing rheumatic carditis [23]. Furthermore, a mutant of *S. mutans*, devoid of expressing antigen I/II yielded on immunisation heart cross-reactive antibody titres similar to those elicited by the antigen I/II intact strain [36]. These findings suggest that antigen I/II does not have heart cross-reactive properties which instead may reside in the cel membrane [37]. Indeed, a M_r 62,000 membrane antigen was identified, having myosin-like epitopes.

Acknowledgements: This work was supported by The Dickinson Trust and by US Public Health Service, Grant R37-DE-08007 to A.S.B. from the National Institute of Dental Research, NIH, Bethesda, MD, USA.

REFERENCES

[1] Loesche, W.J. (1986) Microbiol. Rev. 50, 353–380.
[2] Russell, M.W. and Lehner, T. (1978) Arch. Oral Biol. 23, 7–15
[3] Russell, M.W., Bergmeier, L.A., Zanders, E.D. and Lehner, T. (1980) Infect. Immun. 28, 486–493.
[4] Zanders, E.D. and Lehner, T. (1981) J. Gen. Microbiol. 122, 217–225.
[5] Moro, I. and Russell, M.W. (1983) Infect. Immun. 41, 410–413.
[6] Lehner, T., Russell, M.W., Caldwell, J. and Smith, R. (1981) Infect. Immun. 34, 407–415.
[7] Lehner, T., Caldwell, J. and Avery, J. (1984) Eur. J. Immunol. 14, 814–819.
[8] Russell R.R.B. (1980) J. Gen. Microbiol. 118, 383–388.
[9] Forester, H., Hunter, N. and Knox, K.W. (1983) J. Gen Microbiol. 129, 2779–2788.
[10] Curtiss, R. III. (1985) Curr. Top. Microbiol. Immunol. 118, 253–277.
[11] Bowen, W.H., Cohen, B., Cole, M.F. and Colman, G. (1975) Br. Dent. J. 139, 45–58.
[12] Lehner, T., Challacombe, S.J. and Caldwell, J. (1975) Nature 254, 517–520.
[13] Russell, R.R.B., Beighton, D. and Cohen, B. (1982) Br. Dent. J. 152, 81–84.
[14] McBride, B.C., Song, M., Krasse, B. and Olsson, J. (1984) Infect. Immun. 44, 68–75.
[15] Okahashi, N., Sasakawa, C., Yoshikawa, M., Hamada, S. and Koga T. (1989) Mol. Microbiol. 3, 221–228.
[16] Lee, S.F., Progulske-Fox, A., Erdos, G.W., Piacentini, D.A., Ayakawa, G.Y., Crowley, P.J. and Bleiweis, A.S. (1989) Infect. Immun. in press.
[17] Smith, R., Lehner, T. and Beverley, P.C.L. (1984) Infect. Immun. 46, 168–175.
[18] Lehner, T., Caldwell, J. and Smith, R. (1985) Infect. Immun. 50, 796–799.
[19] Ma, J. K-C., Smith, R. and Lehner, T. (1987) Infect. Immun. 55, 1274–1278.
[20] Ma, J. K-C., Hunjan, M., Smith, R. and Lehner, T. (1989) Clin. Exp. Immunol. in press.
[21] Russell, M.W., Zanders, E.D., Bergmeier, L.A. and Lehner, T. (1980) Infect. Immun. 29, 999–1006.
[22] Lee, S.F., Progulske-Fox, A. and Bleiweis, A.S. (1988) Infect. Immun. 56, 2114–2119.

[23] Okahashi, N., Sasakawa, C., Yoshikawa, M., Hamada, S. and Koga, T. (1989) Mol. Microbiol. 3, 673–678.
[24] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[25] Winship, P.R. (1989) Nucleic Acids Res. 17, 1266.
[26] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich H.A. (1988) Science 239, 487–491.
[27] Mitchell, C.G. and Lehner, T. (1989) Immunology 66, 246–251.
[28] Hollingshead, S.K., Fischetti, V.A. and Scott, J.R. (1986) J. Biol. Chem. 261, 1677–1686.
[29] Takahashi, I., Okahashi, N., Sasakawa, C., Yoshikawa, M., Hamada, S. and Koga, T. (1989) FEBS Lett. 249, 383–388.
[30] Ayakawa, G.Y., Boushell, L.W., Crowley, P.J., Erdos, G.W. McArthur, W.P. and Bleiweis, A.S. (1987) Infect. Immun. 55, 2759–2767.
[31] Ferretti, J.J., Russell, R.B. and Dao, M.L. (1989) Mol. Microbiol. 3, 469–478.
[32] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[33] Van de Rijn, I., Bleiweis, A.S. and Zabriskie, J.B. (1976) J. Dent. Res. 55, c59–c64.
[34] Hughes, M., MacHardy, S.M., Sheppard, A.J. and Woods, N.C. (1980) Infect. Immun. 27, 576–588.
[35] Bergmeier, L.A. and Lehner, T. (1983) Infect. Immun. 40, 1075–1082.
[36] Lee, S.F., Crowley, P.J. and Bleiweis, A.S. (1989) J. Dent. Res. 68, 963. (abstract 773).
[37] Ayakawa, G.Y., Bleiweis, A.S., Crowley, P.J. and Cunningham, M.W. (1988) J. Immunol. 140, 253–257.