Nucleotide Sequence Analysis of the Gene Encoding the *Deinococcus radiodurans* Surface Protein, Derived Amino Acid Sequence, and Complementary Protein Chemical Studies

JÜRGEN PETERS,* MARIETTA PETERS, FRIEDRICH LOTTSPEICH, WOLFRAM SCHÄFER,
AND WOLFGANG BAUMEISTER

Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, Federal Republic of Germany

The complete nucleotide sequence of the gene encoding the surface (hexagonally packed intermediate [HPI])-layer polypeptide of *Deinococcus radiodurans* Sark was determined and found to encode a polypeptide of 1,036 amino acids. Amino acid sequence analysis of about 30% of the residues revealed that the mature polypeptide consists of at least 978 amino acids. The N terminus was blocked to Edman degradation. The results of proteolytic modification of the HPI layer in situ and M_r estimations of the HPI polypeptide expressed in *Escherichia coli* indicated that there is a leader sequence. The N-terminal region contained a very high percentage (29%) of threonine and serine, including a cluster of nine consecutive serine or threonine residues, whereas a stretch near the C terminus was extremely rich in aromatic amino acids (29%). The protein contained at least two disulfide bridges, as well as tightly bound reducing sugars and fatty acids.

There is a growing awareness that regularly arrayed surface proteins (S layers) are widely found in eubacteria and archaebacteria (3, 22, 38). It is not yet clear, however, whether S layers have evolved from a common ancestral gene or whether they have, driven by selection pressure in the natural habitats, evolved independently from other precursor proteins that are capable of reaching the cell surface. The function(s) of bacterial S layers is also still somewhat enigmatic, and it is probably an unreasonable prejudice to search for a common function. A primordial function of perpetual importance in archaebacteria with simple cell envelopes, such as *Thermoproteus tenax*, is shape maintenance (45) and possibly shape determination. S layers, by virtue of their network-like structure, may act as molecular sieves at the interface between the cell and its environment; one aspect of such a sieve is the protection of underlying cell structures from noxious enzymes or exogenous and unsettling DNA. The variability and species specificity of the S-layer outer surface, which is manifested on the level of protein domain structure, led to the hypothesis that S layers can act as mediators of homotypic cell-cell contacts (4).

The structure determination of the surface protein forming the hexagonally packed intermediate (HPI) layer of the radiotolerant bacterium *Deinococcus radiodurans* is particularly advanced. The three-dimensional structure was determined to a resolution of 1.8 nm by electron microscopy (2), and in projection a resolution of 0.8 nm has been attained recently (34). The gene encoding the HPI polypeptide (apparent M_r, 98,000) has been cloned and expressed in *Escherichia coli*. By unilateral deletion with exonuclease III, an ordered set of deletion derivatives of a 5.7-kilobase-pair (kbp) *Hind*III fragment carrying the gene was constructed (33). Here we report the complete nucleotide sequence as well as the results of studies on the amino acid sequence and posttranslational modifications of the HPI polypeptide. The primary structure is indispensable in pursuing our long-range goal of a high-resolution structure determination. It is also expected to shed some light on the phylogenetic relationship

among eubacterial and archaebacterial S layers, a problem that is directly related to the evolution of cell envelopes.

MATERIALS AND METHODS

Chemicals, enzymes, and primer for nucleotide sequencing reactions were obtained from Pharmacia (Freiburg, Federal Republic of Germany [FRG]), except for 7-deaza-dGTP, which was obtained from Boehringer GmbH (Mannheim, FRG). Acrylamide, urea, and sodium dodecyl sulfate (SDS), all electrophoresis grade, were from Bio-Rad Laboratories (Munich, FRG). [α-thio-^{35}S]dATP was purchased from Amersham Corp. (Braunschweig, FRG). Lithium dodecyl sulfate was bought from Serva (Heidelberg, FRG). Trypsin, chymotrypsin, and Tris (Trizma Base) were from Sigma (Munich, FRG). Staphylococcal Glu-C protease and Lys-C protease were obtained from Boehringer. Dimethylaminoazobenzene isothiocyanate and CNBr (sequanal grade) were from Pierce Chemical Co. (Geisenheim-Johannisberg, FRG), and trifluoroacetic acid (TFA; sequence analysis grade) was bought from Fluka (Neu-Ulm, FRG). Hydroxylamine and organic solvents (all chromatography grade) were from E. Merck AG (Darmstadt, FRG).

Oligonucleotide probes. Oligonucleotides (14- to 18mers) were kindly prepared by D. Oesterhelt as described previously (29). As judged by polyacrylamide gel electrophoresis, they were sufficiently pure to be used without further purification, except for desalting on Sephadex-G25 minicolumns.

Nucleotide sequence analysis procedures. Deletion derivatives of the plasmids pJP231 and pJP232 (33) were used as templates in the procedure described by Chen and Seeburg (11). The M13 reverse sequencing primer (Pharmacia) or synthetic oligonucleotides were used. In case artifacts attributable to the secondary structure of the template were encountered, the sequencing reactions were carried out at 50°C, with a concomitant reduction of the incubation period to 5 min. To avoid band compressions, dGTP was either replaced by dITP (30) or 7-deaza-dGTP (31). The radioactive label was [α-thio-^{35}S]dATP (6). Samples were separated on wedge-shaped (0.25 to 0.85 mm), 6% acrylamide gels, and "sharks teeth" combs were used for sample application (11).

---

* Corresponding author.

Preparation of HPI layer. The HPI protein was prepared as described previously (33).

Polypeptide cleavage procedures. CNBr cleavage of the HPI polypeptide was carried out in 70% formic acid with a protein-to-CNBr weight ratio of 1:1. The reaction was stopped by freeze drying. Hydroxylamine cleavage was performed as described previously (7), with an incubation period of 12 h. Specific acid cleavage was conducted with 70% formic acid at 37°C for 24 to 48 h. Proteolytic cleavage with staphylococcal Glu-C protease and with Lys-C protease took place in 50 mM Tris (pH 7.8)–0.5% SDS at 30°C for 10 to 20 h. The HPI polypeptide was denatured at 100°C for 5 min in 2% SDS immediately before digestion. The protease (1:50, by weight) was added in three fractions, with intermittent heating of the digest to 100°C for 5 min. For column chromatography (see below) the samples were extracted twice with 5 volumes of isopentanol-formic acid, which were added to the aqueous phase and precipitate to a final concentration of 70%. Subcleavage with trypsin and chymotrypsin was carried out with SDS-denatured HPI polypeptide after dilution to 0.05% SDS in 50 mM Tris hydrochloride (pH 8.0) for 6 h at room temperature.

Cleavage of the native HPI protein with Glu-C and Lys-C proteases was carried out in 0.1 M NH_4HCO_3 (pH 7.8) containing 20 and 40% (vol/vol) acetonitrile, respectively, at room temperature for 14 h with Glu-C or 48 h with Lys-C. The bulk HPI layer was then sedimented at 11,000 × *g* for 5 min in a centrifuge (Eppendorf), and the pellet was extracted with 0.1 M NH_4HCO_3 containing up to 60% (vol/vol) acetonitrile and finally with chloroform-methanol (2:1; vol/vol). The extracts were taken to dryness in a concentrator (Speed-vac; Bachofer, Reutlingen, FRG).

Peptide separation techniques. Polypeptides with an apparent M_r above 6,000 were separated by SDS-polyacrylamide gel electrophoresis (24) on 5-mm-thick gels. Ten percent of the peptide mixture was labeled with dansyl chloride (27) and used as internal markers. Bands were excised under UV light (366 nm) and electroeluted by using a sample concentrator (Isco) and 100 mM Tris acetate (pH 8.6). Recovered polypeptides were precipitated with 80% (vol/vol) ethanol at −20°C.

Peptides with an apparent M_r below 6,000 were separated by molecular sieve chromatography on 60-cm columns (TSK 2000; LKB Instruments, Inc., Rockville, Md.) in 0.1% TFA–30% acetonitrile or by reversed-phase chromatography on RP columns (Vydac) by using a linear gradient of 0 to 60% acetonitrile in 0.1% TFA. With this solvent system, the effective fractionation range of the TSK 2000 column was ca. 300 to 8,000. Detection was done by determining the A_{206}.

Amino acid analysis. For routine screening of isolated peptides, a micromethod employing TFA-HCl vapor-phase hydrolysis and a precolumn derivatization with *ortho*-pthaldialdehyde with peptide amounts of ca. 0.3 μg gave very good results. A conventional method of hydrolysis with 5.7 M HCl at 110°C for 24 h, separation on an amino acid analyzer (Biotronik), and detection with ninhydrin was also used with ca. 10 μg of protein.

Amino acid sequence analysis. Peptides were routinely screened by using the manual dimethylaminoazobenzene isothiocyanate method (10). For automated sequencing (14), a prototype spinning cup sequenator (26) or a gas-phase sequenator (470A; Applied Biosystems) was used. The phenylthiohydantoin derivatives were analyzed with a high-pressure liquid chromatographic (HPLC) system that separates all components isocratically (25).

Fatty acid analysis. To remove noncovalently bound lipid

material, ca. 10 nmol of the HPI protein was washed three times with 25 mM Tris hydrochloride (pH 7.5)–1% SDS at 60°C and extracted with organic solvent (19). This was followed by denaturation with 25 mM Tris hydrochloride (pH 7.5)–2% SDS at 100°C, precipitation with 4 volumes of ethanol, and size-exclusion chromatography on a TSK 3000 column, with 0.1% TFA–30% acetonitrile used as the solvent. Alternatively, the HPI protein was subjected to preparative SDS-polyacrylamide gel electrophoresis and recovered from the gel by electroelution and precipitated with ethanol. Samples were then treated with methanol in 1 M HCl–methanol in sealed tubes at 100°C for 4 h. Myristic acid was used as an internal standard. Fatty acid analyses were performed with a gas chromatograph and a mass spectrometer plus data system consisting of a gas chromatograph (Fractovap 2101; Carlo Erba, Milan, Italy), a mass spectrophotometer (CH7A; Varian MAT, Bremen, FRG), and a data system (SS200/MS; Finnigan MAT, Bremen, FRG). Separations were performed on fused silica capillaries (30 m by 0.32 mm; DB1 and DB1701; J. and W. Scientific Inc., Ranco Cordova, Calif.) by using helium as the carrier gas, splitless injection, and the temperature program of 2 min at 130°C and then 5°C/min up to 280°C. Alternatively, other columns (50 m by 0.25 mm; CP Sil 88; Chrompack, Middelburg, The Netherlands) and the temperature program 2 min at 100°C and then 2°C/min up to 180°C were used.

Carbohydrate analysis. Sugars were converted to their alditol acetates as described previously (18). To detect amino sugars, samples were first subjected to reductive desamination as described previously (18). For analysis of alditol acetates the gas chromatographic-mass spectrometric-multiple ion detection technique with the ions *m/z* 259 and *m/z* 289 was employed. Separations were performed on a fused silica capillary (DB1701; see fatty acids above) with the temperature program of 2 min at 170°C and then 30 min up to 280°C.

RESULTS

Nucleotide sequence analysis. A set of deletion derivatives generated from the 5.7-kbp *Hin*dIII fragment carrying the *hpi* gene (33) was used in a directed plasmid sequencing approach. Rapid screening (33) of 500 clones and the subsequent mapping of 60 selected plasmids, of which 21 were actually sequenced, proved sufficient for generating an overlapping set of sequences covering a range of 3.4 kbp. The counterstrand was sequenced by using a set of plasmids containing 5'-terminal deletions of the cloned 5.7-kbp *Hin*dIII fragment. Because the deletion method used was found to be satisfactory only for deletions up to about 3.5 kbp, a set of 10 synthetic oligonucleotide sequencing primers was also used for sequencing the counterstrand. Priming sites were selected with the help of the fold algorithm described by Zuker and Stiegler (46) to obviate problems with secondary structures of the template.

The nucleotide sequence of the *hpi* gene is shown in Fig. 1. To confirm the reliability of the sequencing data, a statistical evaluation with the codon preference algorithm (15) was performed (data not shown). One single large open reading frame, ranging from nucleotide positions 1 to 3108, was found (Fig. 1). As many as 74 stop codons were distributed throughout the other two reading frames. The codon usage was distinctly nonrandom throughout the gene. Downstream of the translation termination signal there was no oligo(dT) stretch such as that which is common to most factor-independent terminator regions. There was, however,

```
          -190              -170              -150              -130              -110
CGGTGATGAGTTCCTGGCTGAGCAGGCGTCGGCCACCTCGTGCATCGCCTGCTTGTACTCGGAGACGATCTCATCCTGAACTAACCTATGCTGCATGAACGCTGTACACCGATGGACAGG
                                                        -35                            -10  -----

a
            -70              -50              -30        b        -10        +1        10
AAACTCGGCAACTGGCTCCAGCAGGCACAGCAATCGTTGACGGCGCCGCAGGCGGGATACAGATCGCTTCGCATCCGCTCACGGAGGACTTATGAAGAAAAATATTGCACTCATGGCTCT
                                                                         S.D.      M  K  K  N  I  A  L  M  A  L
                                                                                   1

            50               70         c        90               110              130                      d
CACCGGGGTTTTGACCCTGGCATCCTGCGGTCAGAACGGGAACACCCCGACCGCAGACACTACTGCCCCGACGGTCTCGCTGAGCGTGAACAACGCCAACCTGCCCAGCGGTGTTGGTTC
 T  G  V  L  T  L  A  S  C  G  Q  N  G  N  T  P  T  A  D  T  T  A  P  T  V  S  L  S  V  N  N  A  N  L  P  S  G  V  G  S
            20                                                             40

            170              190              210              230        e  250
TGTCGTTCTGAGCGGCACTGTCAACGAAGCCTCGACGGTCGTTGTTAAAAACAACGCCGGGACGACCGTGTGCACCGTAGAAGTCGCCGCATCCGGCACCTTCACCTGCCCCGCGACGAC
 V  V  L  S  G  T  V  N  E  A  S  T  V  V  V  K  N  N  A  G  T  T  V  C  T  V  E  V  A  A  S  G  T  F  T  C  P  A  T  T
                           60                                              80

     f
            290              310              330              350              370
CATCGCTGGCAACACCAGCACCACCAGCACCAGCACGAGTTATACCGCAACCGCTACCGACGCCGCGAAGAACGTGGGCACCTCCAGCGTCGTGACGGTGAACGTTGCTGGCGTAAGCAA
 I  A  G  N  T  S  T  T  S  T  S  T  S  Y  T  A  T  A  T  D  A  A  K  N  V  G  T  S  S  V  V  T  V  N  V  A  G  V  S  N
                           100                                             120

            410              430              450              470              490
CCCCGCCCCGACCACTGCCGTTCTGACGATTGACCTGGCCGGTGTGAGCAGCGCTCCTATCACCATCAAGGATGCCAACGGCAACGTTGTTCAGGGGTATGACAATGTCACGGTCAACGA
 P  A  P  T  T  A  V  L  T  I  D  L  A  G  V  S  S  A  P  I  T  I  K  D  A  N  G  N  V  V  Q  G  Y  D  N  V  T  V  N  D
                           140                                             160

            530              550              570              590              610
TAACGCGACCATTACTGTTGCCCGTGGCGTCTACACGGTTACCGCTGGTAACGTGAGTGGCTTCAACGGCCCCACGACCAACTTCCGCGTTGACCTGAGCGGTGGCAACCAGACGGTGAC
 N  A  T  I  T  V  A  R  G  V  Y  T  V  T  A  G  N  V  S  G  F  N  G  P  T  T  N  F  R  V  D  L  S  G  G  N  Q  T  V  T
                           180                                             200

            650              670              690              710              730
GTTGAACTACACCCAGGCCGGCACCACCACCCCCACGCCTGTTGGTTCGATCAACATCTTGACGCCTGCAGTGGGTACAAGCGTTACGGGTGGCTCGACCGTTCGCGTGACCTTTGACAA
 L  N  Y  T  Q  A  G  T  T  T  P  T  P  V  G  S  I  N  I  L  T  P  A  V  G  T  S  V  T  G  G  S  T  V  R  V  T  F  D  K
        220                                                            240

            770              790              810              830              850
AGCTAATGAAGTGCAGTGCATGGTTGGCGGCGCCGCCGCAGTCACCGCGCAGGTCGACTCCACGAGCGGCTACTGCGATGTCGTGGTGCCTAACAGCACTGGCAATGTTGTGATCACTGT
 A  N  E  V  Q  C  M  V  G  G  A  A  A  V  T  A  Q  V  D  S  T  S  G  Y  C  D  V  V  V  P  N  S  T  G  N  V  V  I  T  V
                    260                                         280

            890              910              930              950              970
GATGGGTAAGGGTGTCAACGGCCAGACCGTGACGGCAACCCGTAACATTTCCGTGACGCAGGCAGCGGTTTCCTACGGCGTTGTGACCCCGGCGGGTGACCAGGAACTGACCCTGACTAG
 M  G  K  G  V  N  G  Q  T  V  T  A  T  R  N  I  S  V  T  Q  A  A  V  S  Y  G  V  V  T  P  A  G  D  Q  E  L  T  L  T  S
                    300                                         320

            1010             1030             1050             1070             1090
CGAAGGTATCGTGCGCGATGCCGACAGTGGCTGGCGCCGTCTGGGCCAAGGTGTCAGCACCCCCAGCGACCCCAACCTCAACCTTGATATCTACATCAAGGGCACCGTGAACTTCAGCGT
 E  G  I  V  R  D  A  D  S  G  W  R  R  L  G  Q  G  V  S  T  P  S  D  P  N  L  N  L  D  I  Y  I  K  G  T  V  N  F  S  V
              340                                       360

            1130             1150             1170             1190             1210
CAATGCTCCTGCTGGCCAGAAGGTTGAACTGTTCCTCGCCCGCACTACCGGCTCTGACGTCCCCACCAACGATGATATTCAGGCTGGCGACGTGCTGCGCTCGGTGGCCAGCACCAGCGG
 N  A  P  A  G  Q  K  V  E  L  F  L  A  R  T  T  G  S  D  V  P  T  N  D  D  I  Q  A  G  D  V  L  R  S  V  A  S  T  S  G
                    380                                         400

            1250             1270             1290             1310             1330
CACTGAAACCTTCAGCCTCGACAGCCGCCGCCTCGCCGAATTCGACGGCGTGCGCAAGTGGATCGTCGTGCGTATCAACGGCACCCAGGTGACCTACCAGCCCGTGATCGCCGACAACAA
 T  E  T  F  S  L  D  S  R  R  L  A  E  F  D  G  V  R  K  W  I  V  V  R  I  N  G  T  Q  V  T  Y  Q  P  V  I  A  D  N  K
                    420                                         440

            1370             1390             1410             1430             1450
GGGCCCCCAGCAGCCTGATCCCGAACTCAACGGCGTTCAGAATGCTTACTCCAACATCCTGAACAACTACAACAACAGCGGGCTGACCTACGTGCGCGGCCCTGTGAACGTGTTCACCAG
 G  P  Q  Q  P  D  P  E  L  N  G  V  Q  N  A  Y  S  N  I  L  N  N  Y  N  N  S  G  L  T  Y  V  R  G  P  V  N  V  F  T  S
                    460                                         480

            1490             1510             1530             1550             1570
CAACCCCAGCCTGCAGGACCGCGAATTCGGTCAGGCCCCCGTGGGCTCGTCCTTCGTGCAGCGTCGTCCCGCCGGCTTCGAGAGCATTCGGTACTACCTGGTGCCTGAAAGTGCTTTCAA
 N  P  S  L  Q  D  R  F  G  Q  A  P  V  G  S  S  F  V  Q  R  R  P  A  G  F  E  S  I  R  Y  Y  L  V  P  E  S  A  F  N
              500                                       520

            1610             1630             1650             1670             1690
CAACAAGGCTCTGCAGGAGTCCGACGAGATGCTGCGCGCCAAGGCGGTCAAGAGCGTGGCGACCGTGGTCAGCGCTCCTGTCCTCGAACCCGGCACTGTGAAGGCCACTTCGTTCAGCCG
 N  K  A  L  Q  E  S  D  E  M  L  R  A  K  A  V  K  S  V  A  T  V  V  S  A  P  V  L  E  P  G  T  V  K  A  T  S  F  S  R
              540                                       560

            1730             1750             1770             1790             1810
CGTGATCGGCTCGGGTGCTACCAGCACTGTGGCCCCCAAGGCTCTGGACAACGTGACCTACCGCGTCTACGCGATCTCGCGTGACCAGGTGGGCAACGAAACCGCCAGCGCGACCTACGA
 V  I  G  S  G  A  T  S  T  V  A  P  K  A  L  D  N  V  T  Y  R  V  Y  A  I  S  R  D  Q  V  G  N  E  T  A  S  A  T  Y  D
              580                                       600

            1850             1870             1890             1910             1930
CCTGGTCCGTTTCGACAACGTCGGCCCGACCATCACCGGGAGCGTCATCCGCGACACCTCCGACCTGCCCTTCCCCTCGCAGGAACCCGAGCGTTGCCTGAGCGATATCGCCACCATCAG
 L  V  R  F  D  N  V  G  P  T  I  T  G  S  V  I  R  D  T  S  D  L  P  F  P  S  Q  E  P  E  R  C  L  S  D  I  A  T  I  S
              620                                       640

            1970             1990             2010             2030             2050
TCTGGGTGGCATCGCCGACAACGTTGGTGGCGTGGGCCTGAACCCCGGCCAGGGCCTGACCTTCACCCTCGGTGGTCGTCAGATTCAGGCGGGTCAGTTCGACACCAACCAGCTGGCTGA
 L  G  G  I  A  D  N  V  G  G  V  G  L  N  P  G  Q  G  L  T  F  T  L  G  G  R  Q  I  Q  A  G  Q  F  D  T  N  Q  L  A  D
              660                                       680
```

FIG. 1. DNA sequence and derived amino acid sequence of the *hpi* gene of *D. radiodurans* Sark. Partial amino acid sequences confirmed by N-terminal Edman degradation are underlined. A Shine-Dalgarno sequence (S.D.) (37) is also underlined. Palindromes of the terminator region (see text) are indicated with horizontal arrows. Sequences homologous to −10 and −35 regulatory regions in *E. coli* (17) are indicated with dashed lines. Arrows labeled a to f mark the 3' termini of progressive deletions from the 5' end of the 5.7-kbp *Hin*dIII fragment that were used in mapping studies (see Fig. 2).

```
          2090        2110        2130        2150        2170
CCCTGAGTACACCATCGGTTTCAACAGCCTGACCGACGCGCTCGGCAACCCTGTGGTCACTGCCCCGACCAACGCCAAGGTCTACATCGACAACACCGATCCTACGGTCAACTTCAACCG
 P   E   Y   T   I   G   F   N   S   L   T   D   A   L   G   N   P   V   V   T   A   P   T   N   A   K   V   Y   I   D   N   T   D   P   T   V   N   F   N   R
                      700                                                              720

          2210        2230        2250        2270        2290
TGCGGTCATGCAGGGCACCTACGCCAGCGGCGGGCGCGTGTCTGTCGAGAGCGACGCCAGCGACGGCGGCTGCGGTGTGTACGAAACCCGCCTGTTCTGGGACACCGCCAACGGTGTGGT
 A   V   M   Q   G   T   Y   A   S   G   G   R   V   S   V   E   S   D   A   S   D   G   G   C   G   V   Y   E   T   R   L   F   W   D   T   A   N   G   V   V
                  740                                                              760

          2330        2350        2370        2390        2410
TGACGACGCGACCACCACCCCTGCGATCGGCCACCCCGTGCAGTTCGCTCGCCAGCGGGTCACCGACGGTGCCAAGGCCGATAGCCTGAACGCGGGCTGGAACGCCCTGCAGCTGCCCAA
 D   D   A   T   T   T   P   A   I   G   H   P   V   Q   F   A   R   Q   R   V   T   D   G   A   K   A   D   S   L   N   A   G   W   N   A   L   Q   L   P   N
              780                                                              800

          2450        2470        2490        2510        2530
CGGTGCTGGCGCTGTCTACCTCCGTTCGCTGGTCGTTGACCGCGCCGGTAACGCCACCATCTCCACGACCTCCATCGTTGTCAACGCCAAGATTACCAACCAGGCGCGTCCCCTGCTCGG
 G   A   G   A   V   Y   L   R   S   L   V   V   D   R   A   G   N   A   T   I   S   T   T   S   I   V   V   N   A   K   I   T   N   Q   A   R   P   L   L   G
              820                                                              840

          2570        2590        2610        2630        2650
CGGCTTCGACGCCTTCAAGCGCAACGCCAGCGCCCAGTTCGTGGGTGACGACAACGTCATCGCTGGTGTCAACGGCACGGCTGCGACCCCCAACGTCACCGGTAACAGCGCCCTGGATAA
 G   F   D   A   F   K   R   N   A   S   A   Q   F   V   G   D   D   N   V   I   A   G   V   N   G   T   A   A   T   P   N   V   T   G   N   S   A   L   D   N
          860                                                              880

          2690        2710        2730        2750        2770
CATCCTGAGCCTGGATTCGGTGGGCACGCTGACCACCAACGCTTACCTCCCCCGTGGAGCCACCGAAACCGCCATCACCGAGAAGATCCGTAACGTCGGTGCCTATGGCCGCTTCGATGC
 I   L   S   L   D   S   V   G   T   L   T   T   N   A   Y   L   P   R   G   A   T   E   T   A   I   T   E   K   I   R   N   V   G   A   Y   G   R   F   D   A
              900                                                              920

          2810        2830        2850        2870        2890
CACCCAGTGGAACCTCATCCGTGACTACCAGCTGAACACTGACCCCACCCTGCGGAGCGCCTACGTCAACGCGGGCAACCTGGCGAACCAGCGTGGCAACAACTGGCGCATCCGGACCCC
 T   Q   W   N   L   I   R   D   Y   Q   L   N   T   D   P   T   L   R   S   A   Y   V   N   A   G   N   L   A   N   Q   R   G   N   N   W   R   I   R   T   P
              940                                                              960

          2930        2950        2970        2990        3010
CTGGGTTGAACTGGGCAGCAGCGACACCGCCAACACCCAGCAGAAGTTCGACTTCAACAGCGACCTGCTCAACGACTTCTACTACGGCCGCACCTTCGGCAACAACCACAGCGTGAACCT
 W   V   E   L   G   S   S   D   T   A   N   T   Q   Q   K   F   D   F   N   S   D   L   L   N   D   F   Y   Y   G   R   T   F   G   N   N   H   S   V   N   L
              980                                                              1000

          3050        3070        3090        3110        3130
GTTCTCCTACGACCAGTTCAACGGCGTCGTGAGCGACACCGCCGGCGCGTACAGCTTCTACGGCGAAACCGTCCGGAAGTAAGGAAGTACACCTCTAAAAAAGCCTCCCTCCCCAGGGGG
 F   S   Y   D   Q   F   N   G   V   V   S   D   T   A   G   A   Y   S   F   Y   G   E   T   V   R   K   end
              1020                                                  1036

          3170        3190        3210        3230        3250
GCTTCGTTATCCTGGGCCTCTCACCGCCGCAGCCCTTAGCGAGACCCTCTCTCCGCTGTCCTCGCGTTTCCATTCTGCAGTTTCCCATGCAAAGCAACATGATGTTCT
```

a single region containing palindromic sequences (nucleotide positions 3128 to 3153; Fig. 1) which was identified by the terminator algorithm (8).

There was a Shine-Dalgarno sequence GGAGG (37), which is much more conserved in gram-positive bacteria than in *E. coli* (28). The spacing of 4 bp with respect to the ATG codon was unusually short but not unique (35, 41). There was one more potential start codon at nucleotide position 22 but no corresponding upstream Shine-Dalgarno sequence. Moreover, as outlined below, the methionine-1 initiated a typical leader sequence.

Codon usage. The highly nonrandom codon usage of the *hpi* gene (data not shown) exhibited some characteristics that might explain its relatively weak expression in *E. coli* (33) and prove useful in the isolation of other genes of the genus *Deinococcus*. There was a strong preference of G/C over A/T in codon position 3 (80%), whereas the overall G/C content of the gene was 60%. There was also a strong preference of C over G in position three, where C and G were synonymous, except with the codons for valine and leucine. In particular, the predominant usage of CCC for proline (58% of synonymous codons) is unusual even for G/C-rich genes. A notable exception is the gene coding for isopropylmalate dehydrogenase in *Thermus thermophilus* (20), which belongs to the same phylogenetic division as the genus *Deinococcus* (9). In *E. coli* the occurrence of the CCC codon in strongly expressed genes is only about 1% (16).

Amino acid sequence analysis. The denatured HPI polypeptide was cleaved with CNBr, with 70% formic acid at 37°C, and with Lys-C protease or Glu-C protease. The sequences obtained by N-terminal Edman degradation are

indicated in Fig. 1. There was complete agreement between nucleotide and amino acid sequences. The N terminus of the HPI polypeptide was blocked. To isolate an N-terminal peptide, the native HPI layer was treated with Lys-C protease or with Glu-C protease. Only the N-terminal and C-terminal regions of the HPI polypeptide were found to be significantly susceptible to proteolysis, whereas the integrity of the essentially protease-resistant HPI layer was not affected. Proteolytic fragments were separated from the HPI layer by extraction and sedimentation of the layer. Cleavage with Lys-C protease resulted mainly in the release of a peptide that was recovered only in low yield and found to be blocked to Edman degradation. N-terminal sequence analysis of the N-terminally truncated HPI polypeptide revealed that the protein was cleaved at lysine-113 but not at lysine-66. A small fraction of this peptide, which showed a tendency to aggregate, could be recovered from reversed-phase HPLC columns with 25% acetonitrile–0.1% TFA. The identity of this peptide was confirmed by amino acid analysis. The N-terminal peptide contained about 0.7 nmol of palmitoleic acid per mol (other fatty acids were not determined), which was comparable to the data found with the entire HPI polypeptide (see below). Cleavage of lysine-113 released a peptide with an apparent M_r of 6,000, as judged by SDS-polyacrylamide gel electrophoresis. Attempts to subcleave this peptide with staphylococcal Glu-C protease were unsuccessful. When the native HPI layer was treated with Glu-C protease, however, the protein was cleaved at glutamate-77, as identified by N-terminal sequencing of the HPI polypeptide. The small N-terminal peptide could not be isolated by molecular sieve or reversed-phase HPLC, pos-

sibly because it is rendered strongly hydrophobic by bound fatty acid.

Further evidence as to the location of the N terminus of the mature HPI polypeptide was obtained by comparing the apparent M_r values of polypeptides produced from different deletion derivatives of the 5.7-kbp HindIII fragment with the theoretical values calculated for the deleted gene. With increasing 5'-terminal deletion, the apparent M_r of the HPI polypeptide produced in E. coli remained constant down to position d in Fig. 1 and corresponded to those of the purified reference protein and the protein produced from the complete cloned gene (Fig. 2). A slight decrease in apparent M_r was observed when the deletion reached as far as positions e and f in Fig. 1. The apparent M_r of the polypeptide chain encoded by the stretch of the gene reaching from nucleotide position 1 to position d (Fig. 1) was 5,000 (Fig. 2). Using an error margin of ±2,000 for the apparent M_r values (Fig. 2), we conclude that the N terminus of the mature protein probably lies between positions 31 and 59, which is a Glu-C cleavage site identified by amino acid sequence analysis. Because there was no direct evidence for the formation of lacZ' fusion polypeptides, however, we must assume an additional error margin of 1,000.

To confirm the C-terminal sequence of the HPI polypeptide, the native HPI layer was cleaved with hydroxylamine. The C-terminal peptide ranging from positions 1018 to 1036 was obtained in the aqueous supernatant and purified by reversed-phase HPLC. After subcleavage with chymotrypsin, N-terminal sequencing of the C-terminal peptide (positions 1028 to 1036) to the terminal lysine residue confirmed the result of DNA sequence analysis. There was good

TABLE 1. Carbohydrate and fatty acid content of the HPI layer

Component	mol/mol of polypeptide
Carbohydrate	
Glucose	5.60
Galactose	1.05
Mannose	0.23
N-Acetylglucosamine	Tr[a]
Fatty acid	
$C_{14:0}$	Tr
$C_{15:0}$	0.12
$C_{15:1}$	0.20
$C_{16:0}$	0.32
$C_{16:1}$	0.71
$C_{17:0}$	0.10
$C_{17:1}$	0.12
$C_{18:0}$	Tr
$C_{18:1}$	0.12

[a] Tr, Traces.

agreement between the nucleotide sequence-derived amino acid composition data and the amino acid analysis obtained directly (data not shown). To conclude, by determining a total of about 30% of the amino acid sequence of the HPI polypeptide, it was shown, in conjunction with nucleotide sequence analysis, that the mature polypeptide consists of at least 978 amino acids.

Posttranslational modifications. When CNBr cleavage was performed on the unreduced protein and fragments were separated in two dimensions on an SDS-polyacrylamide gel, first in the absence and then in the presence of dithiothreitol, an unreduced fragment with an apparent M_r of 55,000 was cleaved into polypeptides with apparent M_rs of 32,000 and 20,000 under reducing conditions. From N-terminal sequences, apparent M_r values, and amino acid compositions, these fragments were concluded to correspond to CNBr cleavage products ranging from positions 541 to 733 and 734 to 1036 (Fig. 1); the nucleotide sequence-deduced M_r values were 20,122 and 32,454, respectively. Because both polypeptides contained only one single cysteine residue each, cysteine-642 and cysteine-754 were concluded to form a disulfide bridge in the native HPI protein. When the native HPI polypeptide was cleaved with Glu-C protease, a minor site of cleavage in addition to glutamate-59 was found to be glutamate-77. The digested protein was subjected to repeated extraction, as described above, and then treated with dithiothreitol. A peptide was found to be released into the aqueous supernatant, which was identified by N-terminal sequencing and amino acid analysis as the Glu-C peptide ranging from positions 59 to 77. Lys-C cleavage at lysine-113 following Glu-C treatment without reduction released a peptide which could be cleaved with dithiothreitol and yielded peptides with the N-terminal sequences Ala-Ser-Thr and Val-Ala-Ala. It was concluded that cysteine-74 and cysteine-86 are cross-linked in the native HPI protein. The existence of a third disulfide bridge could not be rigorously demonstrated. However, carboxymethylation and subsequent amino acid analyses yielded a value of 6.2 cysteine residues for the reduced protein and 0 cysteine residues for the unreduced protein. Thus, it is likely that cysteine-256 and cysteine-275 are also cross-linked in the native protein.

The results of carbohydrate analyses are shown in Table 1. The data correspond to about six sugar residues per polypeptide chain. For the determination of fatty acids, the native protein was extensively purified in several steps to

deletion up to nucl.pos. in Fig. 1	app. M_r of HPI-polypeptide (10^3) (lane)	DNA-sequence-deduced M_r of unprocessed polypeptide (10^3)
no deletion	98 ± 2 (A)	108
a	98 ± 2 (A)	108
b	98 ± 2 (A)	108
c	98 ± 2 (B)	105 (106)
d	98 ± 2 (B)	103 (104)
e	95 ± 2 (B)	99 (100)
f	94 ± 2 (*)	98 (97)
isolated HPI-polypeptide	98 ± 2	

a   b

FIG. 2. (a) Apparent (app.) M_r of HPI polypeptides produced in E. coli in comparison with DNA sequence-derived data. The apparent M_r was determined by SDS-polyacrylamide gel electrophoresis (33). The corresponding DNA sequence-deduced M_r values are given for polypeptides starting at the respective deletion site indicated in Fig. 1. Values for lac fusion polypeptides are given in parentheses. Note that the apparent M_r of the HPI polypeptide remained constant when the 5'-terminal deletion of the gene reached as far as position d indicated in Fig. 1, corresponding to the shortening of the nascent polypeptide chain by 48 amino acids. The asterisk indicates that this polypeptide was detected only in trace amounts. (b) Partial in vivo proteolytic processing of the HPI polypeptide in E. coli. Protein immunoblots of total E. coli cell protein were stained with a polyclonal antibody directed against the HPI polypeptide and labeled with a fluorescent marker (33). Lane A is representative of the 5'-terminal deletions of the HindIII fragment down to position b in Fig. 1, whereas the pattern in lane B was observed with deletion derivatives that also lacked the region down to position e in Fig. 1.

minimize contamination with noncovalently bound lipid material, as described above (Table 1). The fatty acid composition of the HPI polypeptide roughly reflected the overall fatty acid composition of D. radiodurans (9), except that all saturated fatty acids were present in a higher proportion in relation to the corresponding monounsaturated fatty acids. It was noted above that the purified N-terminal Lys-C peptide contained about 0.7 mol (16:1) fatty acid per mol, which is comparable with the data presented in Table 1. It seems, therefore, that the fatty acids of the HPI polypeptide are located in the N-terminal region of the HPI polypeptide and that they are covalently bound.

DISCUSSION

Expression of the *hpi* gene in *E. coli*. The progressive deletion of the 5.7-kbp HindIII fragment carrying the *hpi* gene with exonuclease III leads to a virtual failure of corresponding clones to produce the polypeptide with an apparent M_r of 98,000 if the length of deletion exceeds about 2.7 kbp from the 5' end (33). We attributed this change to a deletion of an essential part of the control region. The present data reveal, however, that the deletion of this region does not cause the loss of *hpi* gene expression. Even the removal of the translation start region and part of the coding region does not reduce the level of expression, which indicates that the gene is expressed under control of the *lac* promoter in respective clones. It was confirmed by nucleotide sequencing that the corresponding inserts are in frame with the *lacZ'* gene. If, however, the 5'-terminal deletion proceeded beyond the point designated e in Fig. 1, not a single clone was found to show a significant level of *hpi* gene expression, even if the insertion was in frame. This is in marked contrast with the effect of deletion from the 3' end of the gene, where progressive exonuclease III digestion led to the production of similar amounts of increasingly truncated polypeptides which were apparently processed in a way that completely expressed domains were left largely intact (33). It thus appears that the region of the gene which encodes a segment of the N-terminal region of the mature protein contains a sequence that is essential for the efficient expression of the *hpi* gene in *E. coli*.

From the interpretation of the data shown in Fig. 2 (see above), we further conclude that there must be a leader sequence which is also cleaved in *E. coli*. This also provides a plausible explanation for the significant change in the proteolytic fragmentation pattern of the HPI polypeptide that occurs if the 5'-terminal deletion of the gene causes a truncation of the N-terminal sequence of the nascent polypeptide (Fig. 2). The resulting fusion polypeptides would not be expected to be translocated through the cytoplasmic membrane and fold into a native structure. It must be noted, however, that the major portion of the HPI polypeptide is apparently not digested. The putative leader sequence of the HPI polypeptide shows common features with most published leader sequences, namely, a positively charged N-terminal region followed by a hydrophobic core which is relatively rich in leucine (44). These data support the assignment of the start of translation in Fig. 1.

Structural aspects. It has been emphasized previously (21, 38) that S-layer polypeptides contain an abundance of acidic amino acids. It has also been stated that the absence of cysteine is an essential feature of a protein that must cross a rigid cell wall (21). These statements hold for the DNA sequence-derived amino acid sequence of the outer wall protein (OWP) of *Bacillus brevis* which has recently become

available (43). However, whereas Asp and Glu were more abundant than Asn and Gln, respectively, in the OWP protein, the reverse was true for the HPI protein, which also carried only about half as many net negative charges as the OWP protein. Moreover, the HPI protein contained six cysteine residues, of which at least four were involved in disulfide bridges. The HPI layer protein also possessed some other remarkable structural features which were absent in the OWP protein. The N-terminal region (amino acid positions 60 to 250 in Fig. 1) was unusually rich in serine and threonine (29%), and it contained a most unusual cluster of nine consecutive Ser or Thr residues (positions 95 to 103). From the low carbohydrate content of the HPI layer, it follows that only a few Ser and Thr residues may be glycosylated. Furthermore, the Ser and Thr cluster was dispensable for S-layer connectivity, as revealed by the N-terminal proteolytic modification data presented above. The primary structure of the C-terminal region of the HPI layer was also unusual in that it contained, in addition to many polar residues, as much as 29% aromatic amino acids (positions 985 to 1030). It is interesting that a hydrophilic C-terminal region with a high content of aromatic amino acids has been speculated to be engaged in interactions with nucleic acids in the DNA-binding protein of *E. coli* (32). It seems remarkable in this context that electron micrographs of purified HPI-layer preparations consistently show attached DNA strands (unpublished observations).

The low content of reducing sugars found in the HPI polypeptide compares with that of some other eubacterial S-layer polypeptides (21, 39). However, this is, to our knowledge, the first report of an S-layer protein containing fatty acids, and it is unusual for a bacterial protein to contain both fatty acid and carbohydrate. Thus, there are a number of structural features which add to the exceptional properties previously found with *D. radiodurans* (9, 13).

The hydropathy profile (23) of the HPI polypeptide (data not shown) revealed that the C-terminal region is hydrophilic, whereas the N-terminal region is moderately hydrophobic. Results of physicochemical studies have indicated that the HPI layer interacts strongly with the underlying outer membrane via hydrophobic bonding (42). These data are supported by the observation that treatment with SDS at elevated temperatures is required to completely remove the intimately associated outer membrane from the HPI layer. It is interesting that the OWP layer, which is not associated with a membrane and may be removed with chelating or chaotropic agents (21), has a hydrophilic N-terminal region. It seems quite possible that the bound fatty acids and the N-terminal region of the HPI polypeptide serve to anchor the layer to the outer membrane of *D. radiodurans*. Corroborative evidence comes from proteolytic digestion studies which revealed that the native HPI layer is processed by trypsin at lysine-113, lysine-153, arginine-178, and arginine-199 after removal of the outer membrane with SDS but not in the presence of the outer membrane (unpublished data).

Infrared spectroscopic measurements have previously revealed (5) that the HPI layer contains about 30% β structure and virtually no α helix. When the present sequence was analyzed with an algorithm based on the predictive rules described previously (12), 31% of the residues were assigned to the β structure, 7% were α helical, and 11% were turns (considering only strong propensities). The distribution and length of β strands and reverse turns corroborate the hypothesis that most of the β structure is present as an antiparallel β-sheet (data not shown) (5).
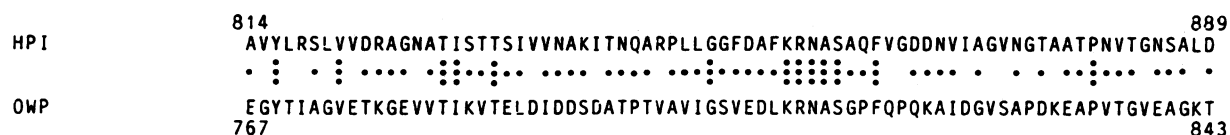
Sequence homologies. A comparison of the amino acid

```
              814                                                                                          889
HP I    AVYLRSLVVDRAGNATISTTSIVVNAKITNQARPLLGGFDAFKRNASAQFVGDDNVIAGVNGTAATPNVTGNSALD
        • :  • :  ••••  •::••:••  ••••    ••••  •••:••••:::::••:   ••••  •   •  • ••:•••  ••• •
OWP     EGYTIAGVETKGEVVTIKVTELDIDDSDATPTVAVIGSVEDLKRNASGPFQPQKAIDGVSAPDKEAPVTGVEAGKT
              767                                                                                          843
```

FIG. 3. Sequence alignments of polypeptide segments. The amino acid sequence of the HPI polypeptide was compared with that of OWP of *B. brevis* (43). The alignment was optimized with respect to the segment comparison score. Identical matches are marked with three dots; conservative replacements, defined as a single-residue comparison score of ≥ 0, in the mutation data matrix (1) are marked with a single dot. The location of each segment within the respective sequence is indicated with amino acid position numbers. The segment comparison score value for the alignment displayed here was 5.1 when the complete sequences were randomized in the calculation and 6.5 if the segments that were compared were randomized.

sequence of the HPI polypeptide with that of the OWP protein of *B. brevis* by use of the relate algorithm (36) revealed that there is a statistically significant homology between two segments of 76 amino acids which were also displaced very little with respect to each other in the two polypeptide chains of almost equal length (Fig. 3). The two segments that were compared might be derived from a common ancestral gene. At this low level of homology, however, more S-layer sequences are required to arrive at conclusions about evolutionary relationships among S-layer polypeptides.

ACKNOWLEDGMENTS

LITERATURE CITED

1. **Barker, W. C., and M. O. Dayhoff.** 1972. Detecting distant relationships: computer methods and results, p. 101–110. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, D.C.

2. **Baumeister, W., M. Barth, R. Hegerl, R. Guckenberger, M. Hahn, and W. O. Saxton.** 1986. Three-dimensional structure of the regular surface layer (HPI-layer) of *Deinococcus radiodurans*. J. Mol. Biol. **187:**241–253.

3. **Baumeister, W., and H. Engelhardt.** 1987. Three-dimensional structures of bacterial S-layers, p. 109–154. *In* J. R. Harris and R. W. Horne (ed.), Electron microscopy of proteins, vol. 6. Membranous structures. Academic Press, Inc., London.

4. **Baumeister, W., and R. Hegerl.** 1986. Can S-layers make bacterial connexons? FEMS Microbiol. Lett. **36:**119–125.

5. **Baumeister, W., F. H. Karrenberg, R. Rachel, A. Engel, B. Ten Heggeler, and W. O. Saxton.** 1982. The major cell envelope protein of *Micrococcus radiodurans* (R1) structural and chemical characterization. Eur. J. Biochem. **125:**535–544.

6. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S-label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80:**3963–3965.

7. **Bornstein, P., and G. Balian.** 1977. Cleavage at asn-gly bonds with hydroxylamine. Methods Enzymol. **47:**132–145.

8. **Brendel, V., and E. N. Tifonov.** 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. **12:**4411–4427.

9. **Brooks, B. W., R. G. E. Murray, J. L. Johnson, E. Stackebrandt, C. R. Woese, and G. E. Fox.** 1980. Red-pigmented micrococci: a basis for taxonomy. Int. J. Syst. Bacteriol. **30:**627–646.

10. **Chang, J. Y., D. Brauer, and B. Wittmann-Liebold.** 1978.

11. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4:**165–170.

12. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47:**46–148.

13. **Counsell, T. J., and R. G. E. Murray.** 1986. Polar lipid profiles of the genus *Deinococcus*. Int. J. Syst. Bacteriol. **36:**202–206.

14. **Edman, P., and G. Begg.** 1967. A protein sequenator. Eur. J. Biochem. **1:**80–91.

15. **Gribskov, M., J. Devreux, and R. R. Burgess.** 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12:**539–549.

16. **Grosjean, H., and W. Fiers.** 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18:**199–209.

17. **Hawley, D. K., and R. McClure.** 1983. Compilation and analysis of *E. coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237–2255.

18. **Henry, R. J., A. B. Blakeney, P. J. Harris, and B. A. Stone.** 1983. Detection of neutral and amino sugars from glycoproteins and polysaccharides as their alditol acetates. J. Chromatogr. **256:**419–427.

19. **Inouye, S., K. Takashi, N. Lee, M. DeMartin, A. Hirashima, and M. Inouye.** 1976. Lipoprotein from the outer membrane of *Escherichia coli*: purification, paracrystallization, and source properties of its free form. J. Bacteriol. **127:**555–563.

20. **Kagawa, Y., H. Nojima, N. Nukiwa, M. Ishizuka, T. Nakajima, T. Yasuhara, T. Tanaka, and T. Oshima.** 1984. High guanine plus cytosine content in the third letter of codons of an extreme thermophile. J. Biol. Chem. **259:**2956–2960.

21. **Koval, S. F., and R. G. E. Murray.** 1984. The isolation of surface array proteins from bacteria. Can. J. Biochem. Cell Biol. **62:**1181–1189.

22. **Koval, S. F., and R. G. E. Murray.** 1986. The superficial protein arrays on bacteria. Microbiol. Sci. **3:**357–360.

23. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.

24. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.

25. **Lottspeich, F.** 1985. Microscale isocratic separation of phenylthiohydantoin amino acid derivatives. J. Chromatogr. **326:**321–327.

26. **Lottspeich, F., J. Kellermann, A. Henschen, G. Rauth, and W. Müller-Esterl.** 1984. Human low-molecular mass kininogen. Eur. J. Biochem. **142:**227–232.

27. **Lubit, B. W.** 1984. Dansylated proteins as internal standards in two-dimensional electrophoresis and Western blot analysis. Electrophoresis **5:**358–361.

28. **McLaughlin, J. R., C. L. Murray, and C. Rabinowitz.** 1981. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β-lactamase gene. J. Biol.

Micro-sequence analysis of peptides and proteins using the 4-N,N'-dimethylaminoazobenzene-4-isothiocyanate/phenylisothiocyanate double coupling method. FEBS Lett. **93:**205–214.

Chem. **256**:11283–11291.

29. **Michel, H., K. A. Weyer, H. Gruenberg, J. Dunger, D. Oersterhelt, and F. Lottspeich.** 1986. The "light" and "medium" subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: isolation of the genes, nucleotide and amino acid sequence. EMBO J. **5**:1149–1158.

30. **Mills, D. R., and F. R. Kramer.** 1979. Structure-independent nucleotide sequence analysis. Proc. Natl. Acad. Sci. USA **76**: 2232–2235.

31. **Mizusawa, S., S. Nishimura, and F. Seela.** 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of a GTP. Nucleic Acids Res. **14**:1319–1324.

32. **Nakayama, N., N. Arai, M. W. Bond, Y. Kaziro, and K.-I. Arai.** 1984. Nucleotide sequence of *dna B* and the primary structure of the *dna B* protein from *Escherichia coli*. J. Biol. Chem. **259**: 97–101.

33. **Peters, J., and W. Baumeister.** 1986. Molecular cloning, expression, and characterization of the gene for the surface (HPI)-layer protein of *Deinococcus radiodurans* in *Escherichia coli*. J. Bacteriol. **167**:1048–1054.

34. **Rachel, R., U. Jakubowski, H. Tietz, R. Hegerl, and W. Baumeister.** 1986. Projected structure of the surface protein of *Deinococcus radiodurans* determined to 8 Å resolution by cryoelectronmicroscopy. Ultramicroscopy **20**:305–316.

35. **Scherer, G. F. E., M. D. Walkinshaw, S. Arnott, and D. J. Morré.** 1980. The ribosome binding sites recognized by *E. coli* ribosomes have regions with signal character in both the leader protein coding segments. Nucleic Acids Res. **8**:3895–3907.

36. **Schwarz, R. M., W. C. Barker, and M. D. Dayhoff.** 1975. A method for detecting sequence homologies of proteins. NBRF report 08710-750215. National Biomedical Research Foundation, Washington, D.C.

37. **Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16 S ribosomal RNA: complementary to nonsense triplets and ribosome linking sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.

38. **Sleytr, U. B., and P. Messner.** 1983. Crystalline surface layers on bacteria. Annu. Rev. Microbiol. **37**:311–339.

39. **Sleytr, U. B., M. Sara, Z. Küpcü, and P. Messner.** 1986. Structural and chemical characterization of S-layers of selected strains of *Bacillus stearothermophilus* and *Desulfatomaculum nigrificans*. Arch. Microbiol. **146**:19–24.

40. **Stackebrandt, E., and C. R. Woese.** 1981. The evolution of prokaryotes, p. 1–31. *In* M. J. Carlile, J. F. Collins, and B. E. B. Moseley (ed.), Molecular and cellular aspects of microbial evolution. Cambridge University Press, Cambridge.

41. **Stormo, G. D., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10**:2971–2996.

42. **Thompson, B. G., R. G. E. Murray, and J. F. Boyce.** 1982. The association of the surface array and the outer membrane of *Deinococcus radiodurans*. Can. J. Microbiol. **28**:1081–1088.

43. **Tsuboi, A., R. Uchihi, R. Tabata, Y. Takahashi, H. Hashiba, T. Sasaki, H. Yamagata, N. Tsukagoshi, and S. Udaka.** 1986. Characterization of the genes coding for two major cell wall protein from protein-producing *Bacillus brevis* 47: complete nucleotide sequence of the outer wall protein gene. J. Bacteriol. **168**:365–373.

44. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5159.

45. **Wildhaber, I., and W. Baumeister.** 1987. The cell envelope of *Thermoproteus tenax*: three-dimensional structure of the surface layer and its role in shape maintenance. EMBO J. **6**: 1475–1480.

46. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.