On the Amino-Acid Sequence of Flagellin from *Bacillus subtilis* 168: Comparison with Other Bacterial Flagellins

(Salmonella/Proteus/cyanogen bromide cleavage/trypptic digestion)

ROBERT J. DELANGE*, JOYCE Y. CHANG*, JOEL H. SHAPER†‡, RAFAEL J. MARTINEZ‡, STANLEY K. KOMATSU*, and ALEXANDER N. GLAZER*

* Department of Biological Chemistry, UCLA School of Medicine, and the ‡ Department of Bacteriology, UCLA, and 1 Molecular Biology Institute, UCLA, Los Angeles, California 90024

Communicated by Emil L. Smith, August 2, 1973

ABSTRACT A partial amino-acid sequence of *Bacillus subtilis* 168 flagellin is presented. The region of unassigned sequence in this 304-residue polypeptide chain spans residues 158–173. Comparison of the 27-residue amino-terminal CNBr peptide of *B. subtilis* 168 flagellin with that derived from the flagellin of the serologically unrelated strain of *B. subtilis*, W23, shows only three conservative substitutions, whereas the 16-residue carboxyl-terminal peptides derived from these flagellins were identical. The comparison of the very limited sequence information available on the flagellins of *Salmonella* and *Proteus* with that on *B. subtilis* indicates homology between these proteins.

Flagellins derived from flagella of varying gross morphology, isolated from a large number of both peritrichously and polarly flagellated bacterial species, possess a number of common features. In each organism, the flagellar filaments are composed of a single type of protein subunit. The molecular weights of these subunits range, in the main, from 33 to 40,000. The amino-acid compositions of numerous highly purified flagellins show an absence of tryptophan and half-cystine, and a low content of aromatic amino acids (1–4).

These similarities may represent the residuum of a common ancestry shared by this class of proteins. Evidence of such an interrelationship may still be discernible in the amino-acid sequences of flagellins obtained from unrelated organisms.

For *Salmonella* (1, 6) and *Bacillus subtilis* (7) flagellins, tryptic peptide mapping has established that changes in the primary structure of flagellin are responsible for the observed wide serological variation in the flagellar antigen in closely related strains of these organisms. Some single amino-acid replacements can lead to alterations in the entire antigenic behavior of the molecule, presumably as a consequence of conformational change; the effects of other replacements are local (6). Flagellin appears to be a particularly interesting macromolecule for the study of the molecular basis of biologically important changes in the antigenic character of a protein.

Single amino-acid substitutions leading to altered flagellar morphology have been reported both for *Salmonella* (8) and for *B. subtilis* 168 (9). In the latter case, a mutation producing flagella lacking the wild-type long period helix appears to be substitution of a valyl for an alanyl residue at a single position in the primary sequence of flagellin (9).

The above considerations, as well as the fact that knowledge of the primary structure of flagellin is a prerequisite to future attempts to determine the conformation of the mono-

TABLE 1. Amino-acid composition of flagellins from *B. subtilis* 168 and W23

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Number of amino-acid residues per molecule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strain 168</td>
</tr>
<tr>
<td>Lysine</td>
<td>16</td>
</tr>
<tr>
<td>Histidine</td>
<td>4</td>
</tr>
<tr>
<td>Arginine</td>
<td>15</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>40</td>
</tr>
<tr>
<td>Threonine</td>
<td>18</td>
</tr>
<tr>
<td>Serine</td>
<td>24</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>42</td>
</tr>
<tr>
<td>Proline</td>
<td>2</td>
</tr>
<tr>
<td>Glycine</td>
<td>19</td>
</tr>
<tr>
<td>Alanine</td>
<td>40</td>
</tr>
<tr>
<td>Valine</td>
<td>14</td>
</tr>
<tr>
<td>Methionine</td>
<td>8</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>24</td>
</tr>
<tr>
<td>Leucine</td>
<td>20</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>1</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>5</td>
</tr>
</tbody>
</table>

Samples (approximately 1 mg of protein) were hydrolyzed in evacuated glass tubes at 110° for 24, 48, and 72 hr with 1.5 ml of 6 N HCl containing 50 μl of 5% phenol in water. Analyses were performed on the Spino automatic amino-acid analyzer, model 120B.

* The number of amino-acid residues per molecule is calculated on the basis of the assumed presence of four histidine residues in each flagellin. The numbers represent the average of determinations performed after 24, 48, and 72 hr of hydrolysis and are given to the nearest whole number except for the serine and threonine values, which were obtained by extrapolation to zero time. No tryptophan was detected spectrophotometrically in either flagellin.
mer and the nature of the contacts involved in the filament assembly process, prompted us to determine the amino-acid sequence of the flagellin of \textit{B. subtilis} 168. We also report a comparison of the sequences at the amino- and carboxy-terminal regions of this flagellin with those derived from the flagellin of an antigenically distinct \textit{B. subtilis} W23.

\textbf{MATERIALS AND METHODS}

\textit{B. subtilis} W23 was grown in minimal medium (10) with 0.5% glucose and 0.1% N-Z Case (pancreatic digest of casein) on a reciprocal shaker at 37°. \textit{B. subtilis} 168 Trp− was grown in the same medium supplemented with 15 \(\mu\)g of \(\alpha\)-tryptophan per ml. Flagella were isolated and purified as previously described (11).

The sequence of \textit{B. subtilis} 168 flagellin has been determined by investigation of the fragments produced by cleavage with cyanogen bromide (12), of the peptides resulting from tryptic digestion of maleylated protein (13), of tryptic peptides derived from the unmodified protein, and last, by utilizing the specific cleavage with \(N\)-bromosuccinimide (14) of flagellin at the single tyrosine residue at position 142 in this polypeptide chain. Sequential Edman degradation of \textit{B. subtilis} 168 flagellin yielded the sequence H\(_2\)N-Met-Arg-Ile-Asn-X-Asn (12), where residue 5 was later identified as His.

\textit{B. subtilis} 168 flagellin contains eight methionine residues. CNBr cleavage yielded eight of the nine theoretically expected fragments (12). The amino-terminal Met residue was cleaved to a very limited extent with CNBr, and the amino-terminal peptide was isolated in part with the methionine still attached. The sum of the eight CNBr peptides accounted for the amino acid composition of the protein (Table 1) (12). Tryptic digestion of the maleyl derivative of flagellin yielded 14 unique peptides which likewise accounted closely for the amino-acid composition of the protein (13).

We were not able to isolate all of the peptides from the tryptic digest of the unmodified protein. Certain of the peptides aggregated in the course of tryptic digestion, and resisted subsequent attempts at solubilization and fractionation. Certain of the tryptic and chymotryptic peptides derived from the 113-residue peptide CNBr-4, as well as from the 103-residue tryptic peptide T-10 (see Fig. 1), behaved in a similar manner. These difficulties have thus far prevented the complete determination of the sequence of residues in

\begin{figure}
\centering
\includegraphics[width=\textwidth]{sequence.png}
\caption{Amino-acid sequence of \textit{B. subtilis} 168 flagellin. Peptides obtained on CNBr cleavage of flagellin and on tryptic cleavage of the maleylated derivative are shown. Areas of unknown sequence are in parentheses.}
\end{figure}
Biochemistry: DeLange et al.

Amino-terminal CNBr peptide

B. subtilis strain W23

H₂N-Met-Arg-Ile-Asn-His-Asn-Ile-Ala-Ala-Leu-Asn-Thr-Leu-

B. subtilis strain 168

H₂N-Met-Arg-Ile-Asn-His-Asn-Ile-Ala-Ala-Leu-Asn-Thr-Leu-

B. subtilis strain W23


B. subtilis strain 168


Carboxyl-terminal CNBr peptide

B. subtilis strain W23

Hsr-Leu-Ala-Gln-Ala-Asn-Gln-Gln-Pro-Gln-

B. subtilis strain 168

Hsr-Leu-Ala-Gln-Ala-Asx-Glx-Glx-Pro-Glx-

B. subtilis strain W23

Asn-Val-Leu-Gln-Leu-Leu-Arg-COOH

B. subtilis strain 168

Asn-Val-Leu-Gln-Leu-Leu-Arg-COOH

RESULTS AND DISCUSSION

Sequence studies on peptides obtained by the approaches described above have led to the partial sequence proposed in Fig. 1. Inspection of the sequence shows the absence of regularities or particularly unusual features in the distribution of the various residues. B. subtilis 168 and W23 flagellins differ totally in antigenic character (4), and significantly in amino-acid composition (Table 1). Comparison of the sequences of the amino- and carboxyl-terminal CNBr peptides derived from the two proteins, representing a total of 44 residues, demonstrated only the carboxyl terminal portion of peptide CNBr-4. The resistance of the insoluble peptides to further degradation by other proteases has limited other possible approaches to the completion of this region of the sequence.

**Fig. 2.** Comparison of amino- and carboxyl-terminal CNBr peptides of B. subtilis W23 and 168. In the amino-acid sequences, underlining and capitalization indicate substitution. Hsr is homoserine.

**Fig. 3.** Comparison of partial sequences of diverse flagellins. Residues identical in flagellins from two bacterial genera are underlined. Double underlining indicates identities in the proteins from three genera.
three conservative substitutions, all in the amino-terminal peptide, as shown in Fig. 2. The carboxyl terminal CNBr peptides were identical.

Only limited information is available on the primary structure of flagellins from other organisms. However, the few comparisons possible are highly suggestive. For example, as shown in Fig. 3 the carboxyl terminal residues of B. subtilis, Salmonella adelaide (15), and Proteus mirabilis (16) flagellins are Leu-Arg, and close homology may well extend considerably beyond these residues. The fragmentary information on the tryptic peptides of S. adelaide (15) and S. typhimurium (17) is strongly suggestive of homologies between the sequences of these proteins and the flagellin of B. subtilis (Fig. 3). The identical tryptic hexapeptide appears both in B. subtilis 168 and S. typhirnurium flagellins (Fig. 3).

The sequence of the amino-terminal 64 residues of B. subtilis 168 flagellin has been presented (18). Accounts of major parts of the present work are available (12, 13). These and other portions of the studies summarized here will be fully described later.

This investigation was aided by Grants GM 11061 from the National Institutes of General Medical Sciences of the United States Public Health Service, and GB 15844 from the National Science Foundation. J.Y.C. was supported by Training Grant GM 00364 from the National Institutes of General Medical Sciences, United States Public Health Service.