Amino Acid Sequences of *Clostridium pasteurianum* Flavodoxin

J. L. Fox and S. S. Smith

Department of Zoology and

J. R. Brown

Department of Chemistry, University of Texas at Austin

(Z. Naturforsch. 27 b, 1096—1100 [1972]; received May 10, 1972)

The amino acid sequences of 10 tryptic peptides from *C. pasteurianum* flavodoxin are determined and ordered. The sequence of two thirds of the protein is presented. New electrophoretic techniques for sequence determination are also presented.

The isolation of phytoflavin from *Anacystis* and the highly homogeneous purification of flavodoxin from *Clostridium* provided a new class of low molecular weight flavoenzyme dehydrogenases which have rapidly become model systems for flavoenzyme reactions. These proteins generally substitute for ferredoxin in its reactions, possess molecular weights in the range of 15,000 to 23,000, have one tightly bound FMN per apoenzyme, exhibit C.D. spectra analogous to other dehydrogenases, possess highly negative redox potentials, and typically exhibit blue semiquinones. A tabulation of the flavodoxins purified to date is given in Table I.

### Materials and Methods

**Cells:** *Clostridium pasteurianum* strain W-5 was grown from potato cultures on a modified Westlake and Wilson media consisting of per liter: 20 mg

---

### Table I. Amino acid compositions of purified flavodoxins.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>12</td>
<td>9</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Tyr</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1-2</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>His</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Asp</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>20</td>
<td>17</td>
<td>13-14</td>
<td>22</td>
<td>28</td>
<td>25</td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td>Thr</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>14</td>
<td>16</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>14</td>
<td>18</td>
<td>17</td>
<td>18</td>
<td>16-17</td>
<td>25</td>
<td>22</td>
<td>22</td>
<td>16</td>
<td>23</td>
<td>26</td>
</tr>
<tr>
<td>Gly</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>19</td>
<td>14</td>
<td>11</td>
<td>19</td>
<td>25</td>
<td>28</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>Ala</td>
<td>14</td>
<td>6</td>
<td>18</td>
<td>18</td>
<td>15</td>
<td>11</td>
<td>15</td>
<td>24</td>
<td>20</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>CyS</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>15</td>
<td>10</td>
<td>13</td>
<td>10</td>
<td>16</td>
<td>5</td>
<td>6</td>
<td>17</td>
<td>15</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Met</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ile</td>
<td>5</td>
<td>15</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td>9</td>
<td>10</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>Leu</td>
<td>12</td>
<td>8</td>
<td>7</td>
<td>13</td>
<td>14</td>
<td>10</td>
<td>12</td>
<td>18</td>
<td>22</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>Tyr</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Phe</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>148</td>
<td>140</td>
<td>138</td>
<td>152</td>
<td>149</td>
<td>122</td>
<td>154</td>
<td>216</td>
<td>210</td>
<td>187</td>
<td>154</td>
</tr>
<tr>
<td>Mol- weight</td>
<td>14600</td>
<td>16050</td>
<td>15000</td>
<td>16000</td>
<td>16000</td>
<td>14500</td>
<td>17000</td>
<td>22000</td>
<td>22800</td>
<td>23000</td>
<td>21000</td>
</tr>
</tbody>
</table>

Requests for reprints should be sent to Dr. J. L. Fox, Department of Zoology, University of Texas, Austin, Texas 78712.
sucrose, 30 mg MgSO\textsubscript{4}, 25 mg CaCl\textsubscript{2}, 3 mg Na\textsubscript{2}MoO\textsubscript{4}, 0.1 mg MnSO\textsubscript{4}, 2.5 \mu g biotin, 3 mg yeast extract, 15.6 g KH\textsubscript{2}PO\textsubscript{4} and 1.5 g K\textsubscript{2}HPO\textsubscript{4}. A final iron content of 0.1 mg/liter and moderate nitrogen bubbling were used. Cells were harvested and frozen from mid log phase yielding 2 g/liter wet weight.

**Enzyme Preparation:** Frozen cells were mechanically ruptured and batch processed essentially as by KNIGHT and HARDY\textsuperscript{2}. Eluant from the DEAE-cellulose column was crystallized from 90\% saturated ammonium sulfate. 0.5 to 1.0 mg flavodoxin/g wet weight cells was obtained and was found to be essentially homogeneous by SDS gel electrophoresis\textsuperscript{15}.

**Enzyme Digestion:** These were performed for 3 hours with 1\% protein weight of proteolytic enzyme.

**Paper Electrophoresis:** a) Diagnostic: Dansyl arginine or sulfone markers\textsuperscript{16} are added to a digest of 4 mg protein and this is applied at 1 mg/cm to washed Whatman 3MM chromatography paper. This is electrophoresed at 3 Kv for 30 min at pH 6.5 in an xylene tank. The paper is dried and fluorescence are marked under a UV lamp. A 1 cm guide strip is cut and stained in cadmium ninhydrin\textsuperscript{17}. Another 1 cm strip is cut, sewn into a second sheet of paper and electrophoresed at pH 1.9 at 3 Kv for 30 min. This paper is also developed in cadmium ninhydrin, the peptides are marked and labelled (see Fig. 1).

![Fig. 1. Two dimensional electrophoretic map of flavodoxin tryptic digest.](image)

Correlation of these two dimensional electrophoreograms with serum albumin peptides has produced a highly useful interpretation\textsuperscript{18}. Peptides possessing the same charges at pH 6.5 and pH 1.9 are found to align themselves in order of decreasing distance along a straight line radiating from the initial origin. Thus, when combined with compositional data, the number of amide derivatives may be identified from charge classes and assistance in establishing stoichiometries is available from molecular weight correlations within a charge class.

The remainder of the pH 6.5 electrophoresis paper is sewn into another paper and electrophoresed at pH 1.9 for 30 min at 3 Kv. This paper is dipped into a trinitrobenzenesulfonate (TNBS) (0.2\% in acetone) solution, dried and developed in pyridine-H\textsubscript{2}O vapors. The lightly yellow stained peptides are identified with the assistance of the corresponding ninhydrin stained paper. The peptides are cut out, eluted with water or pH 6.5 buffer and are analyzed for their amino acid compositions by normal 6 N HCl hydrolysis. The TNP group is removed from the peptide by the acid hydrolysis.

b) Preparative: Larger quantities (typically 40 mg) are purified for sequencing in an analogous manner to the analytical techniques, except the peptides are eluted with the assistance of the ninhydrin guide strips and the TNBS step is omitted.

**Sequence Determination:** The common manual\textsuperscript{19} and automatic Edman\textsuperscript{20} sequencing techniques were used. Thiazolinones were generally hydrolyzed in 6 N HCl at 130 °C for 18 hours and identified as the free amino acid\textsuperscript{21}. Subtractive analyses were used for threonine or serine identification. Samples from the Beckman Sequencer\textsuperscript{20} were often identified also by thin layer chromatography\textsuperscript{20} and gas chromatography\textsuperscript{22}. Results

The tryptic peptide compositions are indicated in Table II. The peptide indicated T-2f contains two peptides. These sequences are presently being obtained from cyanogen bromide peptides more successfully. We believe that this represents the entire protein.

Two partial tryptic splits of flavodoxin are frequently found. Peptide T-5 has a charge of $-1$ at pH 6.5 and $+3$ at pH 1.9 indicative of two basic residues. Subsequent analysis indicated it was composed of T-4 and Tn4 linked by a lysine-glycine bond. Peptide T-6 similarly possessed charges of $-2$ at pH 6.5 and $+3$ at 1.9. Analysis proved T-6 to be Tn2 and T-7 connected by a lysine-glutamic acid bond.

The absence of lysine or arginine in Peptide T-3 indicates its position at the C-terminus. The sequence was found to be Glu-Leu-Val.
Scheme I. Amino Acid Sequence of *Clostridium pasteurianum* Flavodoxin.

```
1 5 10 15 20 25
< T+3 > ← T n1f ← \[→ \]
\< T -5 \> ← \[→ \]
\< T -4 \> ← \[→ \]
\< T -3 \> ← \[→ \]
\< T -2 \> ← \[→ \]
\< T -1 \> ← \[→ \]
```

```

```
26 30 35 40 45 50
\< T -5 \> ← \[→ \]
\[→ \]< T n1f \[→ \]
\< T -1 \> ← \[→ \]
\< T -4 \> ← \[→ \]
\< T -3 \> ← \[→ \]
\< T -2 \> ← \[→ \]
\< T -1 \> ← \[→ \]
```

```
51 55 60 65 70 75
VAL-ALA-PHE-GLY-SER-PRO-SER-MET-GLY-SER-GLU-VAL-GLU,GLX,GLX,MET,PHE-LEU-ASP-VAL-VAL-SER-
\< T -1 \> ← \[→ \]
\< T -4 \> ← \[→ \]
\< T -3 \> ← \[→ \]
\< T -2 \> ← \[→ \]
\< T -1 \> ← \[→ \]
\< T -2 \> ← \[→ \]
\< T -3 \> ← \[→ \]
```

Sequenator analysis was performed on the whole protein through residue 45. Tryptic (T) peptides are indicated by solid lines below the sequence and chymotryptic (C) peptides providing overlapping evidence are indicated by solid lines above the sequence. Some tryptic peptides were digested by thermolysin (Th), chymotrypsin (C), or pronase (Pr), and the resulting peptides are indicated by dashed lines below the parent tryptic peptide. Positive or subtractive Edman is indicated by ← and a blank indicative of serine or threonine by acid regeneration is shown by \[→ \].
Table II. Compositions of tryptic peptides. Fluorescent peptides indicated by f. Ninhydrin stained peptide colors: p-pink, y-yellow, o-orange. Compositions are given on a relative molar ratio basis for a given peptide.

<table>
<thead>
<tr>
<th></th>
<th>-1</th>
<th>-3</th>
<th>-4</th>
<th>-5</th>
<th>-6</th>
<th>-7</th>
<th>n1f</th>
<th>n2</th>
<th>n4</th>
<th>+1</th>
<th>+3</th>
<th>+5</th>
<th>+7</th>
<th>core</th>
<th>-2f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asx</td>
<td>1.8</td>
<td>4.0</td>
<td>2.0</td>
<td>1.9</td>
<td>1.8</td>
<td>0.9</td>
<td>5.2</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>0.8</td>
<td>0.8</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>5.6</td>
<td>1.3</td>
<td>1.2</td>
<td>1.0</td>
<td>1.2</td>
<td>1.3</td>
<td>2.0</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glx</td>
<td>5.0</td>
<td>1.0</td>
<td>2.6</td>
<td>4.2</td>
<td>1.3</td>
<td>1.1</td>
<td>1.2</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>1.6</td>
<td>1.0</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>2.9</td>
<td>1.0</td>
<td>2.4</td>
<td>2.0</td>
<td>0.8</td>
<td>1.0</td>
<td>3.2</td>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>1.6</td>
<td>1.6</td>
<td>2.9</td>
<td>1.0</td>
<td>1.9</td>
<td>1.0</td>
<td>2.1</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>2.7</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.8</td>
<td>2.4</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>0.7</td>
<td>0.5</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>0.6</td>
<td>1.0</td>
<td>0.8</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>1.0</td>
<td>1.0</td>
<td>1.1</td>
<td>1.8</td>
<td>1.7</td>
<td>1.0</td>
<td>2.1</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>1.5</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Try</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>1.0</td>
<td>0.9</td>
<td>1.8</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CySH</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amides</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>pf</td>
<td>p</td>
<td>y</td>
<td>y</td>
<td>o</td>
<td>o</td>
<td>p</td>
<td>pf</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Peptide T-1 is the largest with 34 residues. We have used a combination of chymotryptic, thermolytic and pronase digestions to elucidate the sequence of this peptide. These data are summarized in Scheme I. Some of this has been presented previously.

Manual Edman degradations with positive residue identification or subtractive analysis were used with the tryptic peptides. The results of these studies are shown in Scheme I. The automated Edman sequencer yielded the first 45 residues which were used to overlap these tryptic peptides. Our automated sequence has been confirmed by another laboratory.

Chymotryptic peptides are also used to indicate tryptic peptide overlap as an additional confirmation of our sequence. The presently completed sequence is shown in Scheme I.

The C-terminal sequence was overlapped by the chymotryptic peptide Cn4. A peptic peptide with two lysine residues which could only fit at the end of the T-1 peptide led to the terminal sequence assignment indicated. This leaves an unknown region consisting of the T-2f, T + 5 and T core peptides. These sequences are under investigation. The cysteine which is implicated in the flavin mononucleotide binding is in this unknown region.

The amino acid sequence for the flavodoxin from *Peptostreptococcus elsdenii* has just been published. Comparing 87 residues from our sequence to their 137 residue protein, 33 positions are identical, or about 38%. The tryptic peptide amino acid sequences obtained in our laboratory for the *D. vulgaris* flavodoxin are so dissimilar to the previous two that only the N terminal peptide can be located.

The functional relationship of flavodoxin to ferredoxin raises questions about their structural homologies. The Fitch computerized comparison of our flavodoxin with the sequence of ferredoxin isolated from the same organism was performed. It demonstrated no homology whatsoever. Thus, we have an instance in which there is a convergent evolution of biochemical function from two entirely different sequences. The active sites possess different cofactors for the very same function. It remains to be seen whether any amino acid homology exists, however it is unlikely.

We wish to thank Dr. John Garver, University of Wisconsin and USPHS FR 00226 for permitting us to grow the bacterial cells. Supported by NSFDP, Biomedical Sciences, URI and NSF GB-27464 grants to JLF and USPHS GM 16577 and HD 2899 grants to JRB.

5 S. Mayhew, Biochim. biophysica Acta [Amsterdam] 235, 276 [1971].
14 H. Bothe, private communication.
18 J. R. Brown, Southwest Regional A. C. S. Meeting, 1968, Austin, Texas.
20 P. Edman and G. Begg, European J. Biochem. 1, 80 [1967].
26 M. Dubourdieu and J. L. Fox, unpublished data.