Structural Analysis of the Glycine-rich, Arginine-rich Histone

III. SEQUENCE OF THE AMINO-TERMINAL HALF OF THE MOLECULE CONTAINING THE MODIFIED LYSINE RESIDUES AND THE TOTAL SEQUENCE*

(Received for publication, March 4, 1969)

YOSHTAKA OGAWA,‡ GIANNCARLO QUAGLIAROTTI,§ JOHN JORDAN, CHARLES W. TAYLOR, WESLEY C. STARBUCK,¶ AND HARRIS BUSCH

From the Department of Pharmacology, Baylor University College of Medicine, Houston, Texas 77025

SUMMARY


2. This sequence agrees with that reported recently by DeLange, Fambrough, Smith, and Bonner (J. Biol. Chem., 244, 319 (1969)) except that the arginine residue they placed at position 44 is found in position 40 (and is italicized) in the above sequence.

3. Four clusters of basic residues (underlined) were present in this sequence. The cluster in the amino-terminal portion contained both the modified amino acids e-N-acetyllysine and e-N-methyllysine at positions 16 and 20. The acetyllysine was present in approximately half of the glycine-rich, arginine-rich histone molecules; the methylated lysine position contained mono and dimethyl groups in a ratio of 1:3. This microheterogeneity makes a total of four possible species of the glycine-rich, arginine-rich histone.

* This work was supported in part by the Cancer Center Grant CA 10893, by Grants FR0025902 and FR002564 of the United States Public Health Service, by Grant P-290 from the American Cancer Society, by a grant from the Jane Coffin Childs Fund, and by Grant Q-272 from the Robert A. Welch Foundation. ¶ On leave from Department of Second Surgery, Osaka University Medical School, Osaka, Japan. § Postdoctoral Fellow of the National Cancer Institute (CA-06154-06). ¶ Faculty Research Associate of the American Cancer Society.

The glycine- and arginine-rich histone, first observed (1) and isolated in a highly purified form (2) in this laboratory, was the initial histone selected for analysis of its sequence of amino acids because methods are available for isolation of this GAR histone in large quantities. The GAR histone is the smallest of the known histones and comprises approximately 20% of the total histones (6).

In the previous paper of this series (7), the sequence of the GAR histone for calf thymus was extended to the 53 residues which compose the carboxyl portion of the molecule. In the present study, the sequence of the amino-terminal portion has been elucidated and the total sequence of the GAR histone from calf thymus is presented.

At the same time as these studies were in progress, completely independent studies on the same histone were carried out by DeLange et al. (8) with different enzymatic and chemical approaches to the problem. In the present abbreviated report of our studies, the sequence obtained agrees with that of DeLange et al. (8) with the exception that the arginine residue placed at position 44 is now placed at position 40.

EXPERIMENTAL PROCEDURE

The preparation of the GAR histone in gram quantities was carried out as described earlier (2, 3). The details of the procedures used in this laboratory for chymotryptic and tryptic hydrolysis of the histones or peptides and the subsequent fractionation of the hydrolysates by Sephadex and ion exchange chromatography have been outlined previously (7, 9), as have the procedures used for digestion with amino peptidase M (7), carboxypeptidase A (7, 9), sequential and terminal amino acid analysis (2, 9), amino acid analysis (2, 3, 10, 11), and peptide mapping (2, 9).

All amino acid compositions were calculated after hydrolysis of the peptides for 20 hours. Corrections were made for the destruction of threonine and serine (2, 3, 10). The yields of peptides were calculated (10), using the single residue value of one amino acid in each peptide and were not corrected for loss.

1 The abbreviations used are: GAR, glycine-rich, arginine-rich; dansyl-1-dimethylaminonaphthalene-5-sulfonyl-; Ac, acetyl; Me, methyl.
2 The GAR histone is also known as β3 (3), histone IV (4), and f2al (5).

4387
incurred during the handling procedures used for isolation. The numbers in parentheses following the peptide designation are the residue numbers, and the percentage number following the composition is the yield of the peptide in question. For Edman degradation, the yields (given in parentheses following the step number) are based on the amount of peptide taken at each step and not on the original sample.

**Thermolysin Hydrolysis**
The material to be digested was dissolved in water containing 2 mM calcium chloride to a concentration of 1 mg peptide per ml. The pH was adjusted and maintained at 8.0 with 0.1 M sodium hydroxide using a pH-stat. Thermolysin (courtesy of Dr. Hiroshi Matsubara, Space Sciences Laboratory, University of California, Berkeley, California) dissolved in 0.05 M Tris chloride (pH 8.0) containing 2 mM calcium chloride was added to the sample. The final concentration of thermolysin was 1 to 2% (w/w). The reaction was carried out for 3 hours at 40° and stopped by reducing the pH to approximately 2.0 with the addition of glacial acetic acid.

**Carboxy-terminal Amino Acid Analysis with Carboxypeptidase**
Carboxypeptidase B ( Worthington) used without further treatment, was dissolved in 0.01 M Tris buffer, pH 7.5, at approximately 2.5 mg per ml. For carboxypeptidase B digestion, aliquots of the peptides or the protein were dissolved in Tris buffer, pH 7.5, and digested at 25° or 37° with carboxypeptidase B (10:1 to 25:1, w/w). Aliquots were taken at various time points and added to citrate buffer, pH 2.2 (11) to stop the reaction. The samples were analyzed without further treatment.

**Chromatography on Carboxymethyl Cellulose**
The very basic thermolysin peptides were fractionated on carboxymethyl cellulose (CM-32, Whatman microgranular) using a procedure described previously (12).

**Nomenclature**—The nomenclature of the peptides has been continued as initiated in the previous papers of this series (7, 9).

The tryptic, chymotryptic, and thermolysin peptides have the prefix T, C, and Th, respectively, and are numbered from the carboxyl terminus. Peptides derived from larger peptides are designated, according to the original peptide and the method by which they are obtained, followed by a capital letter, which denotes their arrangement in the sequence (e.g. C2-T3).

**RESULTS**
The amino acid sequence of the amino-terminal half of the GAR histone is shown in Fig. 1 along with the numbered peptides obtained from the enzymatic degradations. Many of these peptides were isolated by procedures reported in previous studies (7, 9). For example, the chymotryptic hydrolysis of the whole GAR histone was fractionated on a column of Sephadex G-25 (7). Peptide C11 (Fig. 1) was isolated from Peak C (see Fig. 1 of Reference 7) and rechromatographed on Aminex A-5 (7); it had the amino acid composition corresponding to the sequence shown in Fig. 1. Peptide C12 was isolated from Peaks A and B (see Fig. 1 of Reference 7) and rechromatographed on Sephadex G-50 (7); the amino acid composition obtained corresponded to the sequence shown in Fig. 1.

The tryptic peptides of the amino-terminal half of the GAR histone were fractionated on Aminex A-5 (Fig. 2). For each of the tryptic peptides, the amino acid composition corresponded to the sequence shown in Fig. 1. The sequences in Fig. 1 were derived by sequence analysis by dansylchloride, Edman degradation, and carboxypeptidase A or B digestion or both.
Fig. 2. Elution pattern obtained by chromatography of a tryptic hydrolysate of the whole GAR histone on a column (1.2 × 50 cm) of Aminex A-5. Temperature, 55°; flow rate, 40 ml per hour. Each fraction contained 4 ml. A stepwise gradient of pyridine-formate buffers and sodium hydroxide was used as indicated.

The presence of asparagine at position 25 adjacent to aspartic acid at position 24 (Fig. 1) was determined by kinetic studies on the release of these two amino acids by amino peptidase M from Peptide T12 over a period from 10 min to 6 hours. Aspartic acid was released first but at 6 hours equimolar concentrations of aspartic acid and asparagine were released.

The presence of acetyl groups blocking the amino-terminal amino acids of the GAR and other histones has been well documented (2, 6, 13, 25). However, the presence of an acetyl group on the ε-nitrogen of lysine residues in the GAR histone has been shown only recently (14, 15) and prompted a search for ε-N-acetyllysine in the primary sequence of the GAR histone. Among the positions suspected was lysine 16 because of the resistance to tryptic hydrolysis of the bond between lysine 16 and arginine 17 in Peptide T4 and the presence of Peptide T12, which had an overlapping amino-terminal sequence with Peptide T23. Aminopeptidase digestion of Peptides T12 and T23 indicated the presence of ε-N-acetyllysine.3 The presence of the ε-N-acetyllysine residue in Peptide T12 was shown by digestion with aminopeptidase M for 11 hours. Aminopeptidase M digestion of T23 produced no ε-N-acetyllysine. After complete digestion by aminopeptidase M or Peptide T12, the recovery of acetylated lysine indicated that 60% of the lysine residue in position 16 was acetylated. The presence of the acetyl group on lysine residue 16 confirms the report of DeLange et al. (8).

Although the presence of ε-N-methyllysine residues in histones has been repeatedly reported (2–4, 6, 16–18), the presence of ε-N-dimethyllysine residues has been found in histones only recently (2, 16, 19, 20). Although the presence of the trimethyl derivative has also been observed in some histone preparations (21), a trimethyl derivative in the GAR histone was not found in this laboratory. Only lysine residue 20 was found to be methylated in the GAR histone. Peptide T18 was eluted at two positions from the Aminex A-5 resin (Fig. 2). The first peak contained ε-N-dimethyllysine and the second peak contained ε-N-monomethyllysine.4 No unmodified lysine was found in Peptide T18. The amount of dimethyllysine in T18 was 3 times the amount of monomethyllysine, confirming the ratio found in the whole histone (2).

Peptide T23 did not react with ninhydrin and had the same amino acid composition shown previously for the amino-terminal tryptic peptide N-Ac-Ser-Gly-Arg- (2, 8, 13). The sequence was confirmed by hydrolysis with carboxypeptidase B and hydrazinolysis of the remaining peptide (8, 13).

The thermolysin hydrolysate of the whole GAR histone was fractionated on Sephadex G-25 (Fig. 3), and rechromatography on Sephadex G-50, Aminex A-5 resin, or carboxymethyl cellulose (Fig. 4) was employed for further purification of the peptides.

3 The authentic ε-N-acetyllysine was provided through the courtesy of Dr. N. L. Benoiton of the Department of Biochemistry, University of Ottawa, Faculty of Medicine, Ottawa, Canada.

4 The analysis for the methylated lysine derivatives was kindly carried out by Mr. R. Keith Wilson.
The thermolysin hydrolysate of Peptide C12 was fractionated on Aminex A-4 resin. The amino acid compositions obtained for the thermolysin peptides corresponded to the sequences shown in Fig. 1, which were determined by Edman degradation and enzymatic digestion.

The presence of glutamine at position 27 was shown by complete hydrolysis of Peptide Th2 with aminopeptidase M for 3 hours. Earlier studies suggested there were 4 amide nitrogens in the GAR histone (2). The asparagine at position 25 and the glutamines at positions 27 and 93 were accounted for in this and previous studies in this laboratory (9). The presence of an amide on the carboxyl of aspartic acid residue number 64 (Fig. 5) in Peptide T7 (7) in the carboxyl half of the GAR histone was shown in a similar manner as reported by DeLange et al. (8). This amide was not reported in the previous study (7).

The isolation and purification of thermolysin Peptide Th2 is described in more detail below.

**Isolation and Structure of Peptides around Arginine 40 (Fig. 1)**

Peptides T14 and T13 which were eluted from Aminex A-5 (Fig. 2) at two different positions had the same amino acid compositions and the same electrophoretic mobility at pH 3.0. However, Peptide T14 was pink and Peptide T13 was brown after staining with cadmium-ninhydrin (9); this result suggested that the amino-terminal amino acid of Peptide T14 was glycine. Glycine was found to be amino-terminal by dansylation. Carboxypeptidase B digestion of Peptide T13 yielded 1 mole each of lysine and arginine per mole of peptide after 30 min. Carboxypeptidase B digestion of Peptide T14 yielded 1 mole of lysine per mole of peptide after 30 min but no arginine was detected. Peptide T14 was resistant to Edman degradation. The results of Edman degradation of T13 were:

<table>
<thead>
<tr>
<th>Peptide T13 (41 to 44)</th>
<th>Arg-Gly-Gly-Val-Lys (57)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composition:</td>
<td>1.0 2.0 1.1 1.0</td>
</tr>
<tr>
<td>Step 1 (96%):</td>
<td>0.3 2.0 1.3 1.0</td>
</tr>
<tr>
<td>Step 2 (96%):</td>
<td>0.3 1.7 1.3 1.0</td>
</tr>
</tbody>
</table>

Arginine was also shown to be the amino-terminal amino acid by dansylation.

<table>
<thead>
<tr>
<th>Peptide T14 (41 to 44)</th>
<th>Gly-Gly-Val-Lys (7%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composition:</td>
<td>2.4 1.2 1.0</td>
</tr>
</tbody>
</table>

Edman degradation:

| Step 1 (88%):          | 1.6 1.2 1.0          |
| Step 2 (65%):          | 1.0 1.0 1.0          |
| Step 3 (24%):          | 0.8 0.3 1.0          |

Edman degradation of T13:

<table>
<thead>
<tr>
<th>Peptide T13 (37 to 39)</th>
<th>Leu-Ala-Arg (25%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composition:</td>
<td>1.0 1.0 1.0</td>
</tr>
</tbody>
</table>

| Edman degradation:     | 0.1 1.0 1.0       |
| Step 2 (65%):          | 0.2 1.0 1.0       |


**20**

**Gly-Lys-Gly-Gly-Ala-Lys (Ac)-Arg-His-Arg-Lys (Me)-**

**30**

**Val-Leu-Arg-Asp-Asn-Ile-Gln-Gly-Ile-Thr-Lys-**

**40**

**Ala-Ile-Arg-Arg-Leu-Ala-Arg-Arg-Gly-Gly-Val-Lys-**

**50**

**Arg-Ile-Ser-Gly-Leu-Ile-Tyr-Glu-Glu-Thr-Arg-Gly-**

**60**

**Val-Leu-Lys-Val-Phe-Leu-Asn-Val-Ile-Arg-Asp-**

**70**

**Ala-Val-Thr-Tyr-Thr-Glu-Asn-Ala-Lys-Arg-Thr-**

**80**

**Val-Thr-Ala-Met-Asp-Val-Val-Tyr-Ala-Leu-Lys-Arg-**

**90**

**Val-Thr-Ala-Met-Asp-Val-Val-Tyr-Ala-Leu-Lys-**

**100**

**Gln-Gly-Arg-Thr-Leu-Tyr-Gly-Phe-Gly-GlyCOOH**.

Fig. 5. The complete amino acid sequence of the GAR histone
Leucine was also shown to be the amino-terminal amino acid by dacylation.

**Peptides T₄₀ and T₄₅ (40 and 45)**—The presence of free arginine in Fig. 2 was shown by amino acid analysis before and after hydrolysis.

**Thermolysin Peptide Th₃**—Peptide Th₃ was isolated by rechromatography of Peak C (Fig. 3) on carboxymethyl cellulose (Fig. 4).

**Peptide Th₃**

<table>
<thead>
<tr>
<th>Amino acid Composition</th>
<th>Th₃</th>
<th>Th₃-Th₄</th>
<th>Th₃-Th₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysine</td>
<td>1.0</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Arginine</td>
<td>3.0</td>
<td>1.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Glycine</td>
<td>2.1</td>
<td></td>
<td>2.3</td>
</tr>
<tr>
<td>Alanine</td>
<td>0.9</td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>Valine</td>
<td>1.0</td>
<td></td>
<td>0.7</td>
</tr>
<tr>
<td>Leucine</td>
<td>0.9</td>
<td></td>
<td>0.6</td>
</tr>
<tr>
<td>Yield (%)</td>
<td>14</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

Carboxypeptidase B digestion for 30 min released 1 mole each of lysine and arginine per mole of peptide.

To characterize further the sequence of Th₃, this peptide was hydrolyzed with thermolysin again for 17 hours. Two peptides were obtained (Th₃-Th₄ and Th₃-Th₅, Table 1) and separated by paper electrophoresis. The hydrolysis products could only be accounted for on the basis of the sequence shown. This sequence has been confirmed by duplication of the degradative and enzymatic studies on Peptide Th₃ isolated from a second thermolysin cleavage of the GAR histone. The sequence of Th₃ could only match that shown in Fig. 1 on the basis of the trypsin and thermolysin peptides described above. This sequence is reviewed further in the “Discussion.”

**DISCUSSION**

The GAR histone is the first histone for which the total sequence (Fig. 5) has been determined. The data obtained in this laboratory provide a total sequence for the GAR histone which, with the exception of 1 residue, is in complete agreement with that obtained independently by DeLange et al. (8). However, the sequence found in this laboratory was obtained with a different approach than that found by DeLange et al. (8). This makes the over-all identity of sequence more striking.

In this laboratory, the GAR histone was cleaved with trypsin, chymotrypsin, and thermolysin, directly. In the study carried out by DeLange et al. (8), the protein was maleylated prior to trypsin hydrolysis. In both studies, however, all of the trypsin peptides were isolated, and they accounted for all the amino acids of the whole GAR histone. Different results were obtained by chymotryptic cleavage. In the study of DeLange et al. (8), the action of chymotrypsin on the GAR histone completely stopped after 20 min, resulting in only a few chymotryptic peptides. However, in the present study, chymotryptic hydrolysis continued for almost 2 hours and 12 peptides were isolated.

Cleavage by thermolysin provided the overlapping peptides of the amino-terminal half of the sequence found in this laboratory. Thermolysin cleaved the peptide bonds on the amino-terminal side of all the leucine, valine, and isoleucine residues with the exception of the valine residue in Peptide Th₃.

Trypsin cleaved all the lysine and arginine bonds with the exception of the bond between lysine and glycine in Peptide Th₅. However, the acetyl or methyl groups on the ε-nitrogen of lysine inhibited the action of trypsin at these positions (23).

Because of the one difference between the sequence obtained by two different laboratories, a summary of the points in this sequence is reviewed. Although the sequence of the GAR histone was essentially completed in this laboratory approximately 3 months ago (24), a preliminary report on the sequence of histone IV (synonymous to the GAR histone) appeared (15) which showed a difference in that arginine 40 of Fig. 5 was reported to be in position 44 (8, 15). Reinvestigation of the sequence in this area of the GAR histone confirmed the positioning of this arginine in position 40.

The sequence of the GAR histone contains several clusters of basic amino acids which may be functional sites for blocking, binding, or initiation reactions:

One of these clusters is of special interest because it contains the modified lysine residues and is symmetrical around histidine.

All of the acetyl and methyl groups are present only on lysines 16 and 20. However, both of these lysines are present in more than one form in the isolated GAR histone. Lysine 16 is acetylated in only 60% of the histone molecules. The di- and mono-methyl derivatives of lysine 20 are present in a 3:1 ratio. Thus, the GAR histone can exist in at least four different forms. One common sequence in different histones has been observed previously, i.e. the amino-terminal Ac-Ser-Gly-Arg sequence of the GAR histone and the AL histone (2, 13). The similarity of the GAR histone in differing cells and species (4, 9, 15, 22) suggests that these areas have general functions rather than tissue specific functions (6).

---

5 In a previous study (2), the GAR histone was reported to contain 93 residues instead of the 102 shown in Fig. 1. The method by which these data were calculated (10) considers both the molecular weight of the protein and the amino acid composition in determining the probable residues of each amino acid. The original molecular weight used for these calculations (2) was low and accounted for part of the difference in the number of residues obtained and those now found. The data now obtained in our laboratory for amino acid composition and sequential analysis confirm the results of DeLange et al. (22) who showed that this histone contains 102 amino acid residues.
Acknowledgments—We wish to acknowledge with gratitude and pleasure the contributions of Dr. P. Sautiere during the early phases of this work and Mr. Keith Wilson for the use of a system developed by him for separating the methylated amino acids.

REFERENCES

17. Murray, K., Biochemistry, 3, 10 (1964).

Yoshitaka Ogawa, Giancarlo Quagliarotti, John Jordan, Charles W. Taylor, Wesley C. Starbuck and Harris Busch

J. Biol. Chem. 1969, 244:4387-4392.

Access the most updated version of this article at http://www.jbc.org/content/244/16/4387

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/244/16/4387.full.html#ref-list-1