J. Biochem. **100**, 359–368 (1986)

Human Spleen Histone H1. Isolation and Amino Acid Sequence of a Main Variant, H1b[1]

Yoshihide OHE, Hiroaki HAYASHI, and Koichi IWAI

Department of Protein Chemistry, Institute of Endocrinology, Gunma University, Maebashi, Gunma 371

Received for publication, March 14, 1986

The complete amino acid sequence of a main variant, H1b, of human spleen histone H1 was determined, following previous determinations of human spleen histones H2B, H2A, H3, and H4. High-performance liquid chromatography on C_8 silica of the H1 fraction yielded the homogeneous H1b subfraction; this variant was estimated to account for 60% of the total of the four H1 variants. The sequence determination was performed with four main fragments, I to IV, obtained by limited chymotryptic digestion of H1b. Together with direct sequencing by automated Edman degradation of fragments II, III, and IV, fragment I, blocked at the N-terminal, and fragment IV, the C-terminal half the H1b molecule, were sequenced after further digestion with staphylococcal protease and others. The four fragments were aligned with three overlapping peptides each derived from chymotryptic partial fragments, I-II and I-II-III, and intact H1b. Carboxypeptidase digestion of intact H1b confirmed the C-terminal sequence of the molecule. Thus, the total sequence of H1b was completely determined; it consists of a total of 218 amino acid residues, has a molecular weight of 21,734 in the unmodified form, and is completely acetylated at the N-terminal serine residue and partially methylated at the lysine residue 25. This sequence is compared with two mammalian somatic H1 sequences.

We have already determined the amino acid sequences of four histone species of human spleen. The first paper described the preparation and fractionation of the whole histone and the sequence determination of the H2B species, which is probably composed of two variants (*1*). The subsequent papers reported the four variant sequences of H2A (*2*), the two variant sequences of H3 (*3*), and the single sequence of H4 (*4*). The main sequences of the four histone species were identical with those of calf thymus, except for one complete substitution (Ser for Cys-96) in H3.

We report here the fractionation of human spleen H1 species into four subfractions and the complete sequence determination of a main variant, H1b. This is the first complete sequencing of a mammalian somatic H1 species. Parts of this study have already been published in the form of abstracts (*5–7*).

[1] This study was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

Abbreviations: CPase, carboxypeptidase; HPLC, high-performance liquid chromatography; ODS, octadecylsilyl; PTH, phenylthiohydantoin; SDS, sodium dodecyl sulfate; SP, sulfopropyl; TFA, trifluoroacetic acid.

EXPERIMENTAL PROCEDURES AND RESULTS

Fractionation of H1 Variants—A fraction of H1 histone was obtained from human spleen whole histone by CM-cellulose chromatography, as described previously (*1*). The fraction was subjected to HPLC on C_8 silica (Merck LiChrosorb RP-8, 5 μm) and four subfractions were obtained, as shown in Fig. 1. Eight chromatographic runs yielded 29.2 mg of the main subfraction, H1b, from 150 mg of the starting fraction. The H1b thus obtained was a homogeneous variant as shown by SDS-gel electrophoresis (*8*) in Fig. 2. Table IS includes the amino acid composition of the H1b preparation. Amino acid analysis, including that of Me-Lys, was carried out with a Hitachi 835 amino acid analyzer, as described previously (*4*).

Limited Chymotryptic Digestion of H1b—The H1b (*ca.* 5 mg, 220 nmol) was digested with α-chymotrypsin (Sigma type I-S, 1/150 w/w) in 0.1 M Na-acetate buffer (pH 5.0, 1 ml) for 45 min at room temperature. The digestion mixture was subjected to HPLC on ODS silica (Bio-Rad Hi-Pore RP-318) and six fragments were obtained, as shown in Fig. 1S and Table IS. Figure 3 shows the strategy for sequencing of these fragments, together with the complete sequence of H1b thus determined.

Except for N-terminal-blocked fragment I, the fragments (II, III, and IV) were directly subjected to automated Edman degradation with a liquid-phase sequence analyzer (JEOL model JAS-47KS) and the resulting PTH-amino acids were analyzed by HPLC, as described previously (*9*) and will be reported elsewhere. Figure 2S-A shows an example of these results. Fragments II and III were thus sequenced completely; the latter sequence was confirmed by its tryptic peptides shown in Table VIS. Fragment IV was sequenced 38% from the N-terminal. Fragments I, I-II, I-II-III, and IV were further digested and sequenced as described below.

Sequencing of Fragment I—Fragment I (*ca.* 50 nmol) was digested with *Staphylococcus aureus* V8 protease (Miles, 1/30 w/w) in 50 mM NH_4HCO_3 (pH 7.8, 0.7 ml) for 22 h at 37°C. Five peptides were obtained from the digestion mixture by HPLC on ODS silica, as shown in Fig. 3S and

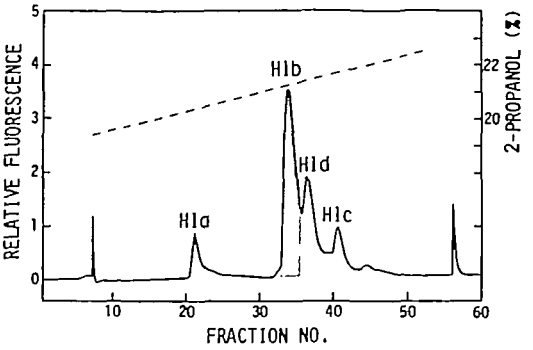


Fig. 1. Chromatography of the starting H1 fraction on a Merck LiChrosorb RP-8 column (8×250 mm) with a 2-propanol gradient in 1.0 M pyridine–formic acid (pH 5.0) at 50°C. Fractions of 4.5 ml/4.5 min were collected, and aliquots were subjected to *o*-phthalaldehyde fluorescence measurement without alkaline hydrolysis.

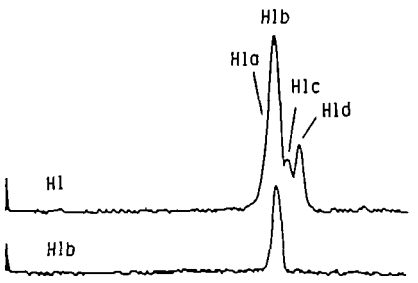


Fig. 2. Electrophoresis of the starting H1 fraction and the H1b subfraction on 15% polyacrylamide slab gel (14×14 cm, 1 mm thick) in 0.1% SDS–0.192 M glycine–25 mM Tris-HCl (pH 8.5) for 9 h at a constant voltage of 150 V. The gel was stained with Coomassie Brilliant Blue R250 and scanned with a microdensitometer at 530 nm.

Table IIS.

Peptide I/S1 and intact H1b were subjected to ^1H NMR measurements, together with calf thymus histone H1, which has *N*-Ac-Ser at the N-terminal (*10*). Figure 4S shows that the three samples have a strong singlet signal at 2.06 ppm, which can be assigned to α-*N*-acetyl protons (*cf.* Fig. 1 in Ref. *11*). The C-terminal Glu of peptide I/S1 was identified by digestion with CPase P (Peptide Institute, Inc.). The digestion mixture also gave a peak, corresponding to that of the authentic *N*-Ac-Ser, in ion-exchange chromatography on Hitachi resin #2618 with 0.1% TFA at 20°C; the peak was shown to yield Ser after hy-

drolysis. Thus, peptide I/S1 was determined to be the N-terminal *N*-Ac-Ser-Glu (Res. 1–2).

Peptides I/S2, S3, S4, and S5 were sequenced by automated Edman degradation; the sequences of Res. 3–9, 16–39, 42–52, and 53–58 were thus determined, respectively. Res. 15 of S2, and Res. 40 and 41 of S3 were determined by CPase P digestion.

For overlapping of peptides I/S1 to S5, fragment I (*ca.* 60 nmol) was digested with thermolysin (Nakarai Chemicals, 1/50 w/w) in 10 mM Tris-HCl, 2 mM $CaCl_2$ (pH 7.0, 0.6 ml) for 2.5 h at 37°C. Four peptides were obtained as shown in Table IIIS. The C-terminal sequence of peptide I/Th1 was determined to be -Glu-(Thr, Ala)-Pro by CPase P digestion. This peptide must be the N-terminal pentapeptide, Ac-Ser-Glu-Thr-Ala-Pro (Res. 1–5), overlapping with peptides I/S1 and S2. Peptides I/Th2, Th3, and Th4 were sequenced by Edman degradation and CPase P digestion. Peptide I/Th2 (Res. 9–18) overlapped with peptides I/S2 and S3, and peptide I/Th4 (Res. 49–55) with peptides I/S4 and S5.

Table IIIS shows that peptide I/Th3 (Res. 23–26) contains Me-Lys, equivalent to about 10% of Lys. Thus, Lys-25 must be partially methylated, which was predicted from the results of Edman degradation of peptide I/S3 and amino acid analysis of the peptide and fragment I.

Overlapping of Fragments I, II, and III—Fragment I-II (*ca.* 40 nmol) was digested with clostripain (Worthington, 1/100 w/w) in 50 mM Tris-HCl, 1 mM $CaCl_2$, 1 mM 2-mercaptoethanol (pH 7.5, 0.2 ml) for 4 h at 37°C. Two peptides were obtained as shown in Table IVS. Peptide I-II/C1 was sequenced by Edman degradation and CPase P digestion; peptide I-II/C2 was sequenced by Edman degradation alone. Thus, peptide I-II/C1 (Res. 33–45) overlapped with peptides I/S3 and S4, and peptide I-II/C2 (Res. 54–63) overlapped with fragments I and II.

Fragment I-II-III (*ca.* 30 nmol) was digested with lysyl endopeptidase (Wako Chemicals, 1/50 w/w) in 20 mM Tris-HCl (pH 9.0, 0.2 ml) for 4 h at 30°C. Peptide I-II-III/L was obtained as shown in Table IVS. This peptide was sequenced by automated Edman degradation using a gas-phase sequence analyzer (Applied Biosystems model 470A). Thus, the sequence of Res. 64–74 overlapped with fragments II and III.

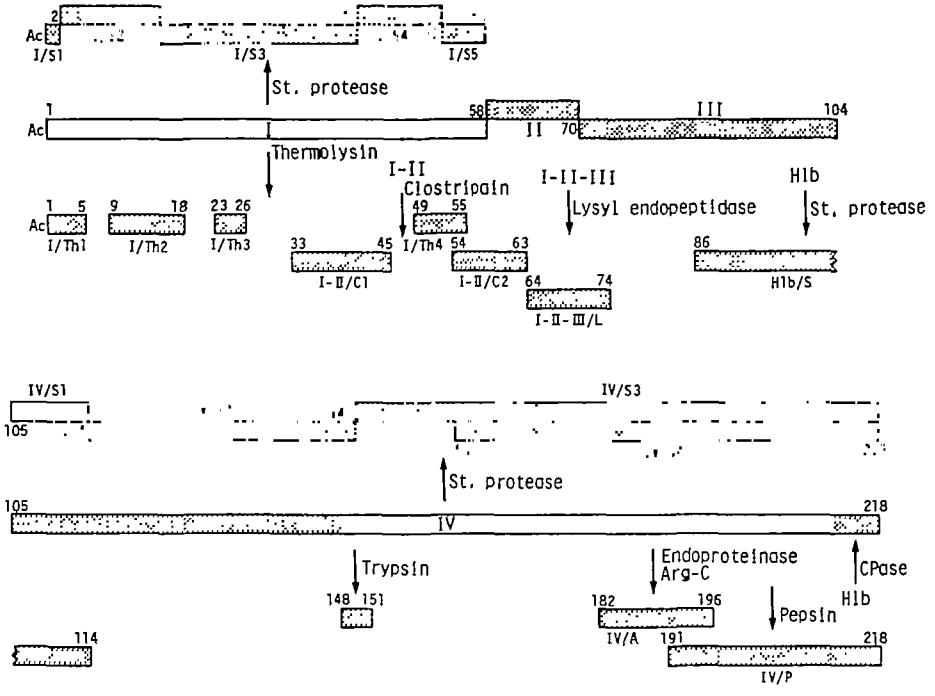
Sequencing of Fragment IV—Fragment IV (*ca.* 80 nmol) was digested with staphylococcal protease (1/30 w/w) in 50 mM NH_4HCO_3 (pH 7.8, 0.2 ml) for 18 h at 37°C. The digestion mixture was subjected to HPLC on ODS silica (Machery-Nagel Nucleosil 300-7C_{18}) and four peptides were obtained, as shown in Fig. 5S and Table VS. Because the sequence of Res. 105–147 had already been determined with fragment IV, peptide IV/S2 (Res. 115–149) was analyzed for the C-terminal sequence by CPase P digestion. Peptides IV/S3 (Res. 150–218) and S4 (Res. 163–218) were sequenced by automated Edman degradation in the liquid phase or gas phase, respectively. The sequences of Res. 142–149, 150–175, and 163–185 were thus determined.

Fragment IV (*ca.* 75 nmol) was digested with endoproteinase Arg-C (Boehringer-Mannheim, 1/50 w/w) in 50 mM $NaHCO_3$ (pH 8.3, 1 ml) for 3 h at 37°C, and peptide IV/A (Res. 182–196) was obtained as shown in Table VS. Fragment IV (*ca.* 100 nmol) was also digested with pepsin (Sigma, 1/20 w/w) in 50 mM acetic acid (pH 3.0, 1 ml) for 3 h at 37°C, and peptide IV/P (Res. 191–218) was obtained (Table VS). These two peptides were sequenced by automated gas-phase Edman degradation (Fig. 2S-B); the sequences of Res. 182–195 and 191–217 were thus determined.

Table VIS summarizes the tryptic peptides obtained from fragment IV. The sequences of these peptides confirmed the results from the above-described sequencing of fragment IV. One peptide (Res. 148–151) also overlapped with peptides IV/S2 and S3.

Total Sequence of H1b—For overlapping of fragments III and IV, intact H1b (*ca.* 50 nmol) was digested with staphylococcal protease (1/30 w/w) in 50 mM NH_4HCO_3 (pH 7.8, 0.5 ml) for 18 h at 37°C, and peptide H1b/S was obtained as shown in Table IVS. This peptide was sequenced by automated gas-phase Edman degradation, and the sequence of Res. 86–114 overlapped with fragments III and IV.

Finally, the C-terminal sequence of intact H1b was determined by digestion with CPase P or CPases A and B (kindly supplied by Dr. T. Tobita, Chiba University), as shown in Fig. 6S. The sequence of -Ala-Ala-Ala-Lys-Lys-Lys (Res. 213–218) thus completed the total sequence of H1b, resulting in a molecular weight of 21,734 in the unmodified form.




```
1                   5                  10                  15                  20                 (Me)
Ac-Ser-Glu-Thr-Ala-Pro-Ala-Ala-Pro-Ala-Ala-Pro-Ala-Pro-Ala-Glu-Lys-Thr-Pro-Val-Lys-Lys-Lys-Ala-Arg-Lys-

                   30                  35                  40                  45                  50
Ser-Ala-Gly-Ala-Ala-Lys-Arg-Lys-Ala-Ser-Gly-Pro-Pro-Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ala-Ala-Ser-

                   55                  60                  65                  70                  75
Lys-Glu-Arg-Ser-Gly-Val-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ala-Leu-Ala-Ala-Ala-Gly-Tyr-Asp-Val-Glu-Lys-Asn-

                   80                  85                  90                  95                 100
Asn-Ser-Arg-Ile-Lys-Leu-Gly-Leu-Lys-Ser-Leu-Val-Ser-Lys-Gly-Thr-Leu-Val-Gln-Thr-Lys-Gly-Thr-Gly-Ala-

                  105                 110                 115                 120                 125
Ser-Gly-Ser-Phe-Lys-Leu-Asn-Lys-Lys-Ala-Ala-Ser-Gly-Glu-Ala-Lys-Pro-Lys-Ala-Lys-Lys-Ala-Gly-Ala-Ala-

                  130                 135                 140                 145                 150
Lys-Ala-Lys-Lys-Pro-Ala-Gly-Ala-Ala-Lys-Lys-Pro-Lys-Lys-Ala-Thr-Gly-Ala-Ala-Thr-Pro-Lys-Lys-Ser-Ala-

                  155                 160                 165                 170                 175
Lys-Lys-Thr-Pro-Lys-Lys-Ala-Lys-Lys-Pro-Ala-Ala-Ala-Ala-Gly-Ala-Lys-Lys-Ala-Lys-Ser-Pro-Lys-Lys-Ala-

                  180                 185                 190                 195                 200
Lys-Ala-Ala-Lys-Pro-Lys-Lys-Ala-Pro-Lys-Ser-Pro-Ala-Lys-Ala-Lys-Ala-Val-Lys-Pro-Lys-Ala-Ala-Lys-Pro-

                  205                 210                 215        218
Lys-Thr-Ala-Lys-Pro-Lys-Ala-Ala-Lys-Pro-Lys-Lys-Ala-Ala-Ala-Lys-Lys-Lys-OH
```

Fig. 3

DISCUSSION

In this study, the chromatographic fractionation of human spleen H1 variants was finally achieved by HPLC on C_8 silica (Fig. 1) and yielded a homogeneous main subfraction, H1b (Fig. 2). This variant was estimated to account for 60% of the total H1 species by analyzing both the chromatographic and electrophoretic peak areas (Figs. 1 and 2).

At the first step of sequencing, limited chymotryptic digestion of H1b, as of H2B (*1*), yielded six fragments including two partially fragmented ones (Fig. 1S and Table IS). The recoveries of these fragments suggest that enzymic cleavage was less than 50% at Leu-58, more than 50% at Tyr-70, and almost complete at Phe-104. All six fragments were very helpful for the complete sequencing of H1b (Fig. 3).

Apart from direct sequencing by automated Edman degradation of fragments II, III, and IV (Fig. 2S), N-terminal-blocked fragment I was further digested with Glu-specific staphylococcal protease and five peptides were obtained with good recoveries (Fig. 3S and Table IIS). However, the recovery of one peptide was about half those of the others; this was due to partial cleavage at Ser-26 by the known nonspecific action of this protease (*12*, *13*). The resulting N-terminal dipeptide, Ac-Ser-Glu, was identified (Fig. 4S) as in the previous cases (*4*, *11*). This and the other four sequenced peptides were aligned with three overlapping thermolysin peptides from fragment I (Table IIIS). Another thermolysin peptide confirmed partial methylation of Lys-25, probably the sole methylated residue in H1b.

The four sequenced fragments, I to IV, were aligned with three overlapping peptides each derived from fragment I-II, fragment I-II-III, and intact H1b (Table IVS). Among the proteases used, Arg-specific clostripain also cleaved at Lys-45 and Lys-63 as known (*3*), lysyl endopeptidase acted specifically, and staphylococcal protease cleaved partially at Ser-85 as expected.

Sequencing of fragment IV was the most difficult in H1b. This fragment covers half the H1b molecule (Res. 105–218) and the N-terminal sequence (Res. 105–147) was directly determined by Edman degradation. Although fragment IV had only one Glu (Res. 114), nonspecific cleavage by staphylococcal protease could be expected as described above. Luckily, the protease also cleaved at Ser-149 more than 50% and at Ala-162 to a much lesser extent, yielding four peptides (Fig. 5S and Table VS). Three of these peptides were used to determine the sequence of Res. 148–185. The remaining C-terminal sequence of 33 residues was determined with two peptides (Res. 182–196 and 191–218) each derived from fragment IV by digestion with endoproteinase Arg-C and pepsin (Table VS); only these two peptides were useful among those obtained with the eight proteases used.

Tryptic peptides obtained from fragments III and IV (Table VIS) were used to confirm each sequencing by Edman degradation and the connection at Ser-149. Finally, the C-terminal sequence determination of intact H1b by using CPases (Fig. 6S) completed the total sequence of H1b, which has a molecular weight of 21,734 in the unmodified form.

Figure 4 shows a comparison of the human spleen H1b sequence with those of calf thymus H1.1 (Res. 1–104, undetermined beyond Res. 105) (*10*) and rabbit thymus H1.3 (the sequence of Res. 1–213 (*14*) was corrected to that of Res. 1–224, as cited (*15*) before publication). The comparison is limited to mammalian somatic H1 species, omitting mammalian testis-specific H1 variants (*15*, *16*) as well as H1 species of lower eukaryotes.

The human sequence differs from the calf and rabbit sequences in 7 and 11 positions (18 and 28% difference) in the N-terminal region of 38/40 residues, and in 0 and 4 positions (0 and 5% difference) in the next internal region of 78 residues, respectively. The human and rabbit se-

Fig. 3. Upper: The method for deduction of the total sequence of H1b. Each bar represents a peptide obtained by limited chymotryptic digestion of H1b (I to IV) and further digestion of I, I-II, I-II-III, IV, and intact H1b with each protease indicated. The shaded parts were sequenced. Lower: The complete sequence of human spleen H1b histone. The N-terminal is completely acetylated and Lys-25 is partially methylated.

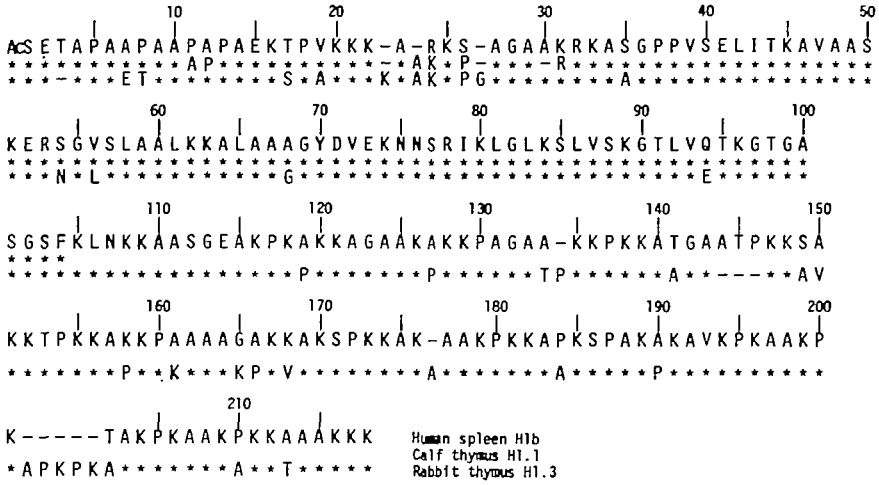


Fig. 4. Comparison of the human spleen H1b sequence with those of calf thymus H1.1 (Res. 1–104 only) (*10*) and rabbit thymus H1.3 (Res. 1–224) (*15*). The sequences are aligned for maximum homology; —, deletion; *, identical with H1b.

quences differ in 26 positions (25% difference) in the remaining C-terminal region of 102/106 residues. Thus, the N- and C-terminal regions are similarly varied, whereas the internal region is strongly conserved among the three mammalian somatic H1 sequences. This internal region of 78 residues contains no Pro and possibly forms the globular structure ('head'), while the N- and C-terminal regions contain many Pro residues and possibly form random coils ('nose' and 'tail'), as proposed (*17*, *18*).

All five histone species of human spleen have now been sequenced completely. The main variant sequences of the four histone species other than H1 are identical with those of calf thymus, except for one complete substitution in H3 (0.7% difference). In contrast, the main variant sequence of human spleen H1 species differs from the calf thymus H1.1 sequence by 7% between Res. 1–104, and from the rabbit thymus H1.3 sequence by 14% between Res. 1–104/106 and 19% between Res. 1–218/224. How such differences in H1 species between mammals correlate with those between human spleen H1 variants will be reported elsewhere.

We thank Prof. T. Miyazawa and Dr. S. Yokoyama of the Univ. of Tokyo for NMR measurements, and the staff of the Applied Bioscience Laboratory, Kirin Brewery Co. and the J.S.I. Laboratories, Japan Scientific Instrument Co. for making available a gas-phase sequence analyzer.

REFERENCES

1. Ohe, Y., Hayashi, H., & Iwai, K. (1979) *J. Biochem.* **85**, 615–624
2. Hayashi, T., Ohe, Y., Hayashi, H., & Iwai, K. (1980) *J. Biochem.* **88**, 27–34
3. Ohe, Y. & Iwai, K. (1981) *J. Biochem.* **90**, 1205–1211
4. Hayashi, T., Ohe, Y., Hayashi, H., & Iwai, K. (1982) *J. Biochem.* **92**, 1995–2000
5. Ohe, Y., Hayashi, H., & Iwai, K. (1983) *Seikagaku* (in Japanese) **55**, 641
6. Ohe, Y., Hayashi, H., & Iwai, K. (1984) *Seikagaku* (in Japanese) **56**, 671
7. Ohe, Y., Hayashi, H., & Iwai, K. (1985) *Seikagaku* (in Japanese) **57**, 753
8. Laemmli, U.K. (1970) *Nature* **227**, 680–685
9. Hayashi, H., Ohe, Y., Hayashi, T., & Iwai, K. (1985) *J. Biochem.* **97**, 441–448
10. Liao, L.W. & Cole, R.D. (1981) *J. Biol. Chem.* **256**, 3024–3029
11. Fusauchi, Y. & Iwai, K. (1984) *J. Biochem.* **95**, 147–154
12. Walker, J.M., Hastings, J.R.B., & Johns, E.W. (1977) *Eur. J. Biochem.* **76**, 461–468
13. Walker, J.M., Goodwin, G.H., & Johns, E.W. (1979) *FEBS Lett.* **100**, 394–398
14. Cole, R.D. (1977) in *The Molecular Biology of the Mammalian Genetic Apparatus* (Ts'o, P., ed.) Vol. 1, pp. 93–104, Elsevier/North-Holland, Amsterdam

15. Cole, K.D., York, R.G., & Kistler, W.S. (1984) *J. Biol. Chem.* **259**, 13695–13702
16. Cole, K.D., York, R.G., & Kistler, W.S. (1986) *Biochim. Biophys. Acta* **869**, 223–229
17. Bradbury, E.M., Cary, P.D., Chapman, G.E., Crane-Robinson, C., Danby, S.E., Rattle, H.W.E., Boublik, M., Palau, J., & Aviles, F.J. (1975) *Eur. J. Biochem.* **52**, 605–613
18. Hartman, P.G., Chapman, G.E., Moss, T., & Bradbury, E.M. (1977) *Eur. J. Biochem.* **77**, 45–51

Supplemental Materials

TABLE IS. Amino acid compositions (molar ratios) and recoveries (mol %) of human spleen Hlb and its chymotryptic fragments (Fig. 1S). Values were obtained on 24 h hydrolysates at 110°C with 6M HCl and corrections were made for losses of Thr and Ser during hydrolysis; values for Hlb are averages of three analyses. Values in parentheses are the numbers of residues in the finally determined sequence (Fig. 3).

Amino acid	Intact Hlb 1–218	Fragment, Res. No. I 1–68	II 59–70	I-II 1–70	III 71–104	I-II-III 1–104	IV 105–218
Asx	4.1 (4)	0.3 (0)	0.1 (0)	0.3 (0)	3.0 (3)	3.2 (3)	1.1 (1)
Thr	9.9(10)	3.0 (3)	0.1 (0)	3.1 (3)	2.9 (3)	6.6 (6)	4.0 (4)
Ser	15.8(16)	7.1 (7)	0.3 (0)	7.3 (7)	5.0 (5)	12.3(12)	4.3 (4)
Glx	6.9 (7)	4.1 (4)	0.2 (0)	4.2 (4)	2.0 (2)	6.3 (6)	1.2 (1)
Pro	20.8(21)	6.7 (7)	0.0 (0)	6.9 (7)	0.0 (0)	7.1 (7)	14.2(14)
Gly	14.1(14)	3.2 (3)	1.3 (1)	4.3 (4)	4.9 (5)	9.2 (9)	5.2 (5)
Ala	59.7(59)	14.7(15)	5.7 (6)	20.3(21)	1.0 (1)	21.6(22)	36.5(37)
Cys	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Val	8.4 (8)	4.0 (4)	0.1 (0)	3.8 (4)	3.0 (3)	6.8 (7)	1.0 (1)
Met	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Ile	1.9 (2)	1.1 (1)	0.0 (0)	1.0 (1)	1.0 (1)	1.8 (2)	0.0 (0)
Leu	9.3 (9)	2.2 (2)	2.0 (2)	4.1 (4)	4.0 (4)	8.0 (8)	1.1 (1)
Tyr	1.1 (1)	0.1 (0)	1.0 (1)	1.0 (1)	0.2 (0)	1.1 (1)	0.0 (0)
Phe	1.0 (1)	0.3 (0)	0.0 (0)	0.0 (0)	1.2 (1)	1.3 (1)	0.0 (0)
Lys	60.8(62)	8.8 (9)	2.2 (2)	11.0(11)	4.9 (5)	15.9(16)	45.4(46)
His	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Arg	4.3 (4)	3.2 (3)	0.1 (0)	2.9 (3)	1.0 (1)	3.9 (4)	0.2 (0)
Recovery	—	31	26	29	53	13	76

TABLE IIS. Amino acid compositions (molar ratios) and final recoveries (mol %) of the peptides obtained by staphylococcal protease digestion of fragment I (Fig. 3S). Values are as in Table IS.

Amino acid	Peptide, Res. No. I/S1 1–2	I/S2 3–15	I/S3 16–41	I/S4 42–52	I/S5 53–68
Asx	0.1 (0)	0.1 (0)	0.1 (0)	0.1 (0)	0.1 (0)
Thr	0.0 (0)	1.0 (1)	1.1 (1)	1.0 (1)	0.0 (0)
Ser	1.0 (1)	0.2 (0)	3.1 (3)	1.1 (1)	1.9 (2)
Glx	1.0 (1)	1.1 (1)	1.1 (1)	1.0 (1)	0.1 (0)
Pro	0.0 (0)	4.0 (4)	2.8 (3)	0.0 (0)	0.0 (0)
Gly	0.1 (0)	0.2 (0)	2.1 (2)	0.1 (0)	1.0 (1)
Ala	0.0 (0)	6.9 (7)	4.9 (5)	2.9 (3)	0.0 (0)
Val	0.1 (0)	0.1 (0)	2.1 (2)	1.1 (1)	1.0 (1)
Ile	0.1 (0)	0.1 (0)	0.2 (0)	1.0 (1)	0.1 (0)
Leu	0.1 (0)	0.1 (0)	0.1 (0)	1.0 (1)	1.0 (1)
Tyr	0.1 (0)	0.1 (0)	0.1 (0)	0.1 (0)	0.1 (0)
Phe	0.1 (0)	0.1 (0)	0.2 (0)	0.0 (0)	0.1 (0)
Lys	0.1 (0)	0.3 (0)	6.7 (7)	2.0 (2)	0.1 (0)
Arg	0.0 (0)	0.1 (0)	2.1 (2)	0.0 (0)	1.0 (1)
Recovery	87	75	36	88	86

TABLE IIIS. Amino acid compositions (molar ratios) and final recoveries (mol %) of the peptides obtained by thermolysin digestion of fragment I. The peptides were separated by HPLC on Nucleosil 300-7C_{18} as in Fig 5S; I/Th2 and Th3 were rechromatographed on Nucleosil 5C_{18} or CM-cellulose, respectively. Values are as in Table IS.

Amino acid	Peptide, Res. No. I/Th1 1–8	I/Th2 9–18	I/Th3 23–26	I/Th4 49–55
Asx	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Thr	0.9 (1)	0.9 (1)	0.0 (0)	0.0 (0)
Ser	0.9 (1)	0.0 (0)	0.9 (1)	1.9 (2)
Glx	1.0 (1)	1.1 (1)	0.0 (0)	1.0 (1)
Pro	1.0 (1)	2.9 (3)	0.0 (0)	0.0 (0)
Gly	0.2 (0)	0.1 (0)	0.1 (0)	0.8 (1)
Ala	1.2 (1)	4.1 (4)	1.0 (1)	1.0 (1)
Val	0.1 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Ile	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Leu	0.1 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Tyr	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Phe	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Lys	0.1 (0)	1.0 (1)	1.0 (1)[a]	1.2 (1)
Arg	0.0 (0)	0.0 (0)	1.1 (1)	1.1 (1)
Recovery	33	32	63	67

[a] Composed of 0.9 Lys and 0.1 Me-Lys

TABLE IVS. Amino acid compositions (molar ratios) and final recoveries (mol %) of the peptides obtained by clostripain digestion of fragment I-II, lysyl endopeptidase digestion of fragment I-II-III, and staphylococcal protease digestion of intact Hlb. The peptides from I-II and I-II-III were separated by HPLC on Nucleosil 5C_{18}, and that from Hlb on Nucleosil 300-7C_{18}, as in Fig 5S. Values are as in Table IS.

Amino acid	Peptide, Res. No. I-II/C1 33–45	I-II/C2 54–63	I-II-III/L 64–74	Hlb/S 86–114
Asx	0.0 (0)	0.0 (0)	1.1 (1)	1.1 (1)
Thr	0.9 (1)	0.0 (0)	0.1 (0)	2.8 (3)
Ser	2.1 (2)	2.0 (2)	0.2 (0)	3.8 (4)
Glx	0.9 (1)	0.0 (0)	1.2 (1)	2.2 (2)
Pro	1.9 (2)	0.0 (0)	0.1 (0)	0.0 (0)
Gly	1.0 (1)	1.0 (1)	1.1 (1)	5.0 (5)
Ala	1.1 (1)	1.9 (2)	3.6 (4)	3.1 (3)
Val	1.0 (1)	1.0 (1)	1.0 (1)	2.0 (2)
Ile	0.9 (1)	0.0 (0)	0.1 (0)	0.1 (0)
Leu	1.1 (1)	2.0 (2)	1.0 (1)	3.1 (3)
Tyr	0.0 (0)	0.0 (0)	1.0 (1)	0.0 (0)
Phe	0.0 (0)	0.0 (0)	0.2 (0)	1.0 (1)
Lys	2.1 (2)	2.0 (2)	1.0 (1)	5.0 (5)
Arg	0.0 (0)	0.0 (0)	0.1 (0)	0.1 (0)
Recovery	78	52	95	50

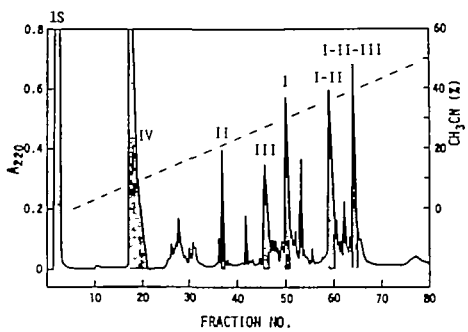
TABLE VS. Amino acid compositions (molar ratios) and final recoveries (mol %) of the peptides obtained from fragment IV by staphylococcal protease, endoproteinase Arg-C, or pepsin digestion. The first two digests were fractionated by RPLC on Nucleosil 300-7C_{18}, and the last one on Nucleosil 3C_{18}, as in Fig. 8S. Values are as in Table IS.

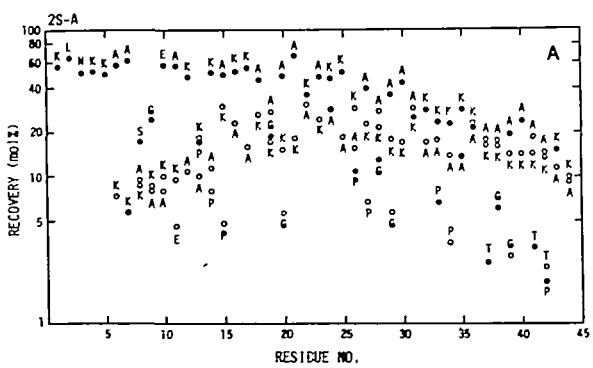
Amino acid	Peptide, Res. No.					
	IV/S1 105–114	IV/S2 115–149	IV/S3 150–218	IV/S4 163–218	IV/A 182–196	IV/P 191–218
Asx	1.0 (1)	0.1 (0)	0.1 (0)	0.0 (0)	0.1 (0)	0.0 (0)
Thr	0.0 (0)	1.9 (2)	1.7 (2)	0.9 (1)	0 1 (0)	1 0 (1)
Ser	1.0 (1)	1.0 (1)	1 8 (2)	1 9 (2)	1 0 (1)	0.1 (0)
Glx	1.0 (1)	0.1 (0)	0.1 (0)	0 0 (0)	0.1 (0)	0.1 (0)
Pro	0.0 (0)	4.0 (4)	10.5(10)	8.1 (8)	2.9 (3)	4.0 (4)
Gly	1.0 (1)	3.0 (3)	1.1 (1)	0.9 (1)	0.2 (0)	0.1 (0)
Ala	2 0 (2)	11.8(12)	23.8(23)	18.9(19)	4.3 (4)	9.1 (9)
Val	0 0 (0)	0.0 (0)	1.1 (1)	1 1 (1)	0.9 (1)	1.1 (1)
Ile	0 0 (0)	0 0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.1 (0)
Leu	1 0 (1)	0 1 (0)	0.1 (0)	0 0 (0)	0 1 (0)	0.0 (0)
Tyr	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Phe	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0 0 (0)	0.0 (0)
Lys	3.0 (3)	13.3(13)	29.1(30)	24.3(24)	6.0 (6)	12.8(13)
Arg	0.0 (0)	0.0 (0)	0.1 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Recovery	70	65	42	6	11	41

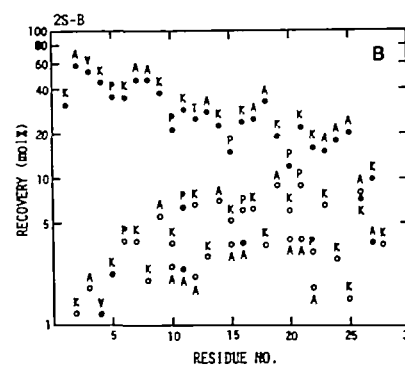
TABLE VIS. Tryptic peptides obtained from fragments III and IV. Each fragment (ca. 40 nmol) was digested with trypsin (Sigma type I, 1/70 (III) or 1/100 (IV) w/w) in 0.1 M $NaHCO_3$ for 1 h (III) or 2.5 h (IV) at 37°C. The digest of III was fractionated by RPLC as in Fig. 1S; that of IV was first chromatographed on SP-Sephadex and then on Nucleosil $5C_{18}$ as in Fig 5S, and further purified on Ritachi resin (#2619, if necessary. Residue numbers are those of the finally determined sequence (Fig 3) Sequences were determined by Edman degradation and CPase P digestion. Recoveries (mol %) are final ones

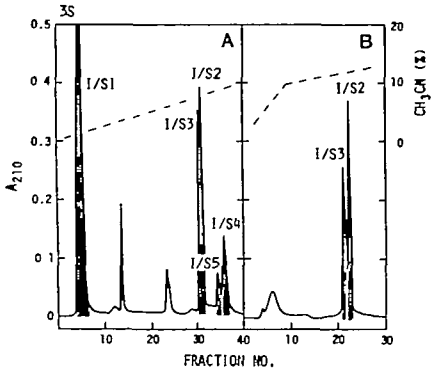
Res. No	Sequence	Recovery
71–74	(Asp,Val,Glu,Lys)	53
71–78	(Asp,Val,Glu,Lys,Ala,Asn,Ser,Arg)	27
75–78	Ala-Asn-Ser-Arg	59
79–80	(Ile,Lys)	36
79–84	(Ile,Lys)Leu-Gly-Leu-Lys	20
81–84	(Leu,Gly,Leu,Lys)	62
85–89	Ser-Leu-Val-Ser-Lys	87
90–96	Gly-Thr-Leu-Val-Glu-Thr-Lys	77
97–104	Gly-Thr-Gly-Ala-Ser-Gly-Ser-Phe	84
105–108	Lys-Leu-Asn-Lys	34
105–109	Lys-Leu-Asn-Lys-Lys	47
109	Lys	135/8[a]
110–118	Ala-Ala-Ser-Gly-Glu-Ala-Lys-Pro-Lys	61
119–120	Ala-Lys	174/6[b]
121	Lys	135/8
121–125	Lys-Ala-Gly-Ala-Ala-Lys	48
122–126	Ala-Gly-Ala-Ala-Lys	14
127–128	Ala-Lys	174/6
127–135	Ala-Lys-Lys-Pro-Ala-Gly-Ala-Ala-Lys	37
129–135	Lys-Pro-Ala-Gly-Ala-Ala-Lys	44
136–138	Lys-Pro-Lys	41
139–147	Lys-Ala-Thr-Gly-Ala-Ala-Thr-Pro-Lys	74
148	Lys	135/8
148–151	Lys-Ser-Ala-Lys	13
149–151	Ser-Ala-Lys	23
152–155	Lys-Thr-Pro-Lys	64
156	Lys	135/8
157–158	Ala-Lys	174/8
159–167	Lys-Pro-Ala-Ala-Ala-Ala-Gly-Ala-Lys	80
168	Lys	135/8
169–170	Ala-Lys	174/6
171–173	Ser-Pro-Lys	30
174	Lys	135/8
175–176	Ala-Lys	174/6
177–181	Ala-Ala-Lys-Pro-Lys	247/3[c]
182–185	Lys-Ala-Pro-Lys	58
186–189	Ser-Pro-Ala-Lys	82
190–191	Ala-Lys	174/6
192–196	Ala-Val-Lys-Pro-Lys	87
197–201	Ala-Ala-Lys-Pro-Lys	247/3
202–206	Thr-Ala-Lys-Pro-Lys	94
207–211	Ala-Ala-Lys-Pro-Lys	247/3
212–216	Lys-Ala-Ala-Ala-Lys	49
217	Lys	135/8
218	Lys	135/8

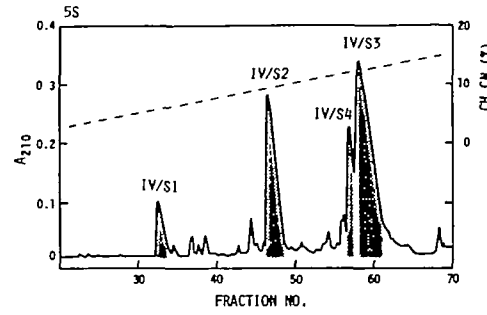
[a] Derived from 8 positions, [b] from 6 positions, [c] from 3 positions.

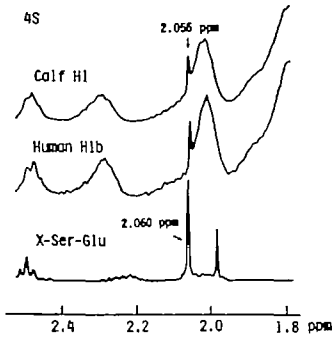












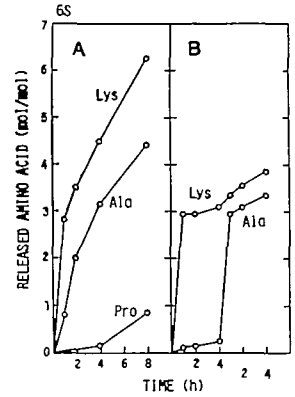


Fig. 1S. Chromatography of the limited chymotryptic digest of Hlb on a Bio-Rad Hi-Pore RP-318 column (4.6 x 250 mm) with an acetonitrile gradient in 0.1 % (v/v) TFA at 50°C. Fractions, 2.0 ml/2.0 min.

Fig. 2S. Recoveries of PTH-amino acids on automated Edman degradation. A. Fragment IV (ca. 20 nmol) was analyzed with a liquid-phase sequence analyzer (JEOL model JAS-47KS). B. Peptide IV/P (ca. 2 nmol) was analyzed with a gas-phase sequence analyzer (Applied Biosystems model 470A). ●, an amino acid found at that residue; ○, those derived from the preceding residues.

Fig. 3S. A. Chromatography of the staphylococcal protease digest of fragment I, as in Fig. 1S. Fractions, 1.0 ml/1.0 min. B. Rechromatography of the fraction of I/S2 and I/S3 on a Hitachi gel #3063 column (4 x 300 mm) with an acetonitrile gradient in 0.1 % (v/v) TFA at 50°C. Fractions, 1.0 ml/1.0 min.

Fig. 4S. ^{1}H NMR spectra of the N-terminal-blocked dipeptide (X-Ser-Glu), intact Hlb, and calf thymus HI in 10 mM DCl-D_2O, recorded on a Bruker WH-400 spectrometer.

Fig. 5S. Chromatography of the staphylococcal protease digest of fragment IV on a Nucleosil 300-7C_{18} column (4 x 300 mm) with an acetonitrile gradient in 0.1 % (v/v) TFA at 40°C. Fractions, 1.0 ml/1.0 min.

Fig. 6S. Carboxypeptidase digestion of intact Hlb. A. Hlb (ca. 8 nmol) was incubated with CPase P (10 μg) in 0.1 M acetic acid (pH 2.9, 0.4 ml) at 30°C. B. Hlb (ca. 10 nmol) was incubated with CPase B (1/50 w/w) in 20 mM Tris-HCl (pH 8.0, 0.6 ml) at 37°C. After 4 h, the digestion mixture was boiled for 5 min at 110°C and then incubated with CPase A (1/50 w/w) at 37°C. Aliquots (0.1 ml) of these digestion mixtures were analyzed at the indicated times.