Amino Acid Sequence of the Human Haptoglobin β Chain. I. Amino- and Carboxyl-Terminal Sequences†

Don R. Barnett,* Tong-Ho Lee,‡ and Barbara H. Bowman

ABSTRACT: The amino acid sequence of the heavy (β) chain of human haptoglobin has been investigated. The sequence of the first 36 residues of the amino-terminal region and the sequence of the last nine residues of the carboxyl-terminal region are presented. These sequences were determined on tryptic, chymotryptic, and cyanogen bromide fragments. The first 24 residues of the intact, reduced, and alkylated β chain were determined on an automatic sequencer. Residues 19–24 provided an overlap with CNBr fragment II. These studies demonstrate unexpected structural homologies between the N- and C-terminal sequences of the haptoglobin β chain and the corresponding regions of bovine chymotrypsin A and bovine trypsin. Residues of the haptoglobin β chain which are identical with either bovine chymotrypsin A or trypsin are in italic type. *Ile*-Leu-*Gly*-*Gly*-His-Leu-Asp-Ala-Lys-*Gly*-*Ser*-Phe-*Pro*-*Trp*-*Gln*-Ala-Gly-Met-Val-Leu-Gly-Gly-His-Leu-Asp-Ala-Lys-Ala-Glu-Leu-Ile-Asn-Gln-Gln-Trp-Leu. The C-terminal nonapeptide sequence is *Trp*-*Val*-*Glx*-Lys-*Thr*-*Ile*-*Ala*-Glu-*Asn*. The haptoglobin β chain also displays internal homology within the N-terminal region. The sequence of residues 2–9 is identical with that of residues 20–27.

Human haptoglobin (Hp)[1] is an α_2-globulin which demonstrates the physiological function of specific and stoichiometric binding of free hemoglobin in human plasma (Polonovski and Jayle, 1938; Jayle and Moretti, 1962; Nagel, 1967) and is also implicated in the normal catabolism of red blood cells by the reticuloendothelial system (Wada *et al.*, 1970). The monomeric, functional unit of Hp is composed of four polypeptide chains, two α chains and two β chains (Shim and Bearn, 1964). The α chains provide the genetic polymorphism observed in Hp (Smithies *et al.*, 1962a; Black and Dixon, 1968). The α-2 chains are responsible for the multiple polymer formation of Hp 2-1 and Hp 2-2. The Hp β chain was shown by electrophoresis in acid–urea gels (Smithies *et al.*, 1962b) to be identical in all common genetic types. Confirmation was provided by immunological studies (Shim *et al.*, 1965) and by the demonstration of identical tryptic peptide maps and amino acid compositions (Cleve *et al.*, 1967). The Hp β chain is important in the hemoglobin-binding function of the haptoglobin molecule (Gordon and Bearn, 1966). This polypeptide is comprised of some 280 amino acid residues and contains the carbohydrate moieties of the glycoprotein (Cleve *et al.*, 1969).

It has been observed during the present study of the amino acid sequence of the Hp β chain that an internal homology exists between the first 9 residues and residues 19–27. Furthermore, when the C- and N-terminal sequences of the Hp β chain were compared to sequences of other proteins, an unexpected observation was made. The C- and N-terminal sequences of the Hp β chain show rather extensive similarities between the corresponding C- and N-terminal sequences of the serine proteinases, *e.g.*, bovine chymotrypsin A and bovine trypsin.

In this, the first of a series of papers reporting the amino acid sequence of the Hp β chain, experiments leading to the determination of the first 36 residues from the amino terminus and the 9 C-terminal residues are presented.

Materials and Methods

Haptoglobin and β-Chain Purification. Haptoglobin was purified from human plasma or ascites fluid by the method of Gordon and Bearn (1966). The β-polypeptide chain was isolated by the method of Gordon *et al.* (1968) with the following modifications. The molar concentrations of β-mercaptoethanol and iodoacetamide were increased twofold to achieve complete reduction and alkylation. Reduced and alkylated α and β chains were separated on a 2.5 × 62 cm column packed with Sephadex G-200 equilibrated with 5 M guanidine hydrochloride–0.05 M Tris-HCl (pH 7.75) (Figure 1). The G-200 had been sieved and only those particles 53–74 μ in diameter were used. The OD of the column effluent was monitored at 280 mμ by a LKB Uvicord system. Purity of the β-chain preparations was established by observing a single band on acid–acrylamide gel electrophoresis (Jordan and Raymond, 1969) of a 10-mg/ml protein solution.

Carboxyl-Terminal Labeling. The intact Hp β chain was selectively labeled on the C-terminal residue by reaction with [^{3}H]water (Matsuo *et al.*, 1966; Holcomb *et al.*, 1968). Identification of the C-terminal residue was made after acid and enzymatic (carboxypeptidase) hydrolysis of the ^{3}H-labeled Hp β chain followed by separation of amino acids on a Beckman 116 amino acid analyzer. The effluent was collected in 1 ml fractions and counted on a liquid scintillation counter.

Cyanogen Bromide Fragment Preparation. Hp β chain (1 μmole) containing 1 mg of tritiated β chain was subjected to cleavage with CNBr (Gross, 1967). The reaction was carried out in 5 ml of 70% formic acid for 72 hr at room temperature using 0.5 g of CNBr (Pierce Chemical Co.). The fragments were separated by chromatography on a 2 × 100 cm Sephadex

† From the Department of Human Genetics, University of Texas Medical Branch, Galveston, Texas 77550. *Received September 30, 1971.* This work was supported in part by grants from The Robert A. Welch Foundation (H-378), The National Institutes of Health (HD 03321), and The Harris and Eliza Kempner Fund.

‡ Present address: Catholic Medical College, Seoul, Korea.

[1] Abbreviations used are: Hp, human haptoglobin; 4-SPITC, 4-sulfophenyl isothiocyanate; 4-SPTH, 4-sulfophenylthiohydantoin; DMAA, dimethylallylamine; DNS, 1-dimethylaminonaphthalene-5-sulfonyl chloride; PTH, phenylthiohydantoin; Quadrol, [*N*,*N*,*N*′,*N*′-tetrakis(2-hydroxypropyl)ethylenediaminetrifluoroacetic acid]; PITC, phenyl isothiocyanate.

TABLE I: Amino Acid Composition of the Hp β Chain, N-Terminal Cyanogen Bromide Fragment (CNBr I), the Second Cyanogen Bromide Fragment (CNBr II), the C-Terminal Cyanogen Bromide Fragment (CNBr VI), the N-Terminal Tryptic Peptide (T-5-E), the N-Terminal Chymotryptic Peptide (C-14-B), and the Second Chymotryptic Peptide (C-9-B).

Amino Acid[a]	Hp β	CNBr 1	CNBr II	CNBr VI	T-5-E	C-14-B	C-9-B
Lysine	22.0 (22)	1.0 (1)	4.2 (4)	1.0 (1)	1.0 (1)		1.1 (1)
Histidine	10.0 (10)	0.9 (1)	2.3 (2)		0.9 (1)	1.1 (1)	
Arginine	4.9 (5)		0.7 (1)				
Aspartic acid	28.4 (28)	1.2 (1)	5.1 (5)	1.1 (1)	1.3 (1)		1.1 (1)
Threonine	18.8 (19)		3.8 (4)	1.0 (1)			
Serine	14.5 (15)	0.8 (1)	2.2 (2)				1.0 (1)
Glutamic acid	29.1 (29)	1.0 (1)	5.1 (5)	2.1 (2)			
Proline	12.8 (13)	0.8 (1)	1.7 (2)				0.8 (1)
Glycine	20.3 (20)	3.8 (4)	2.5 (3)		2.1 (2)	1.9 (2)	1.1 (1)
Alanine	20.9 (21)	1.9 (2)	3.1 (3)	1.1 (1)	1.1 (1)		1.0 (1)
Half-cystine as CMC[d]	5.5 (6)						
Valine	26.6 (27)		4.5 (5)	1.0 (1)			
Methionine or homoserine	3.7 (4)[e]	0.6 (1)	0.7 (1)				
Isoleucine	13.1 (13)	1.0 (1)	2.4 (2)	0.9 (1)	0.9 (1)	0.9 (1)	
Leucine	23.4 (23)	1.8 (2)	6.0 (6)		1.9 (2)	2.0 (2)	
Tyrosine	11.5 (12)		1.3 (1)				
Phenylalanine	8.0 (8)	0.9 (1)	1.1 (1)				1.0 (1)
Tryptophan[b]	(5)[c]	+ (1)	+ (1)	+ (1)			+ (1)
Total	280	18	48	9	9	6	8

[a] Values are given in residues per mole. The assumed integral values are given in parentheses. [b] Trp Ehrlich stain. [c] Five tryptic peptides contained Trp. [d] CMC-carboxymethylcysteine. [e] One residue of Met is presumed destroyed in acid hydrolysis.

G-100 column equilibrated with 30% acetic acid. A flow rate of 6 ml/hr was maintained by a Beckman-Spinco pump. The column effluent was monitored on an automatic peptide analyzer (Jones, 1966). The C- and N-terminal fragments were eluted together in the last peak (Figure 2). Subsequent separation was achieved by ion-exchange chromatography using Bio-Rad Aminex A-5 resin on an automatic peptide analyzer following the procedure of Jones (1966). The C-terminal CNBr peptide was identified by the ^{3}H label on the C-terminal residue. The N-terminal CNBr fragment, fragment I, was identified by the presence of N-terminal isoleucine, which is the N-terminal residue of the Hp β chain.

Tryptic and Chymotryptic Peptide Preparation. Hp β chain (1 μmole) was subjected to enzymatic hydrolysis with trypsin and 1 μmole with chymotrypsin. Trypsin (three-times crystallized) and α-chymotrypsin (three-times crystallized) were obtained from Worthington Biochemical Corp. Digestions were carried out on a pH-Stat (Radiometer) using a 100:1 protein:enzyme ratio. The pH, during digestion, was automatically maintained at pH 8.0 by addition of 1% ammonium hydroxide. Digestion was carried out for 5 hr at 37°. The digests were taken to dryness on a flash evaporator and dissolved in 0.3 M pyridine–HCl (pH 2.0) for peptide analysis and separation on an automatic peptide analyzer using ion-exchange chromatography on Bio-Rad Aminex A-5 and AG 50W-X2 resins according to the method of Jones (1966; R. T. Jones, 1969, personal communication). The detailed description of the separation and analysis of tryptic and chymotryptic peptides obtained from Hp β chain will be described in subsequent papers. The amino acid composition of each peptide, after acid hydrolysis in 6 N HCl for 18 hr, was determined on a Beckman Model 116 amino acid analyzer. Table I gives the composition of peptides included in this study.

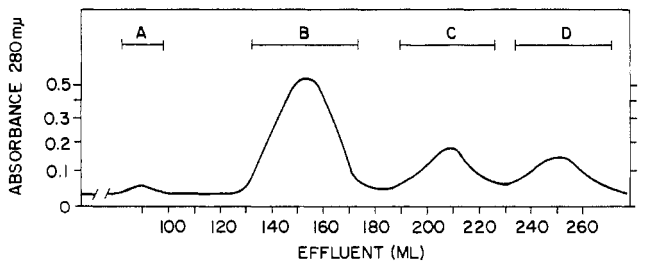


FIGURE 1: Elution profile of reduced and alkylated Hp 2-1 on Sephadex G-200 (2.5 × 62 cm) equilibrated in 5 M guanidine hydrochloride–0.05 M Tris-HCl (pH 7.75). (A) Unreduced Hp, (B) Hp β chain, (C) Hp α-2 chain, and (D) Hp α-1 chain.

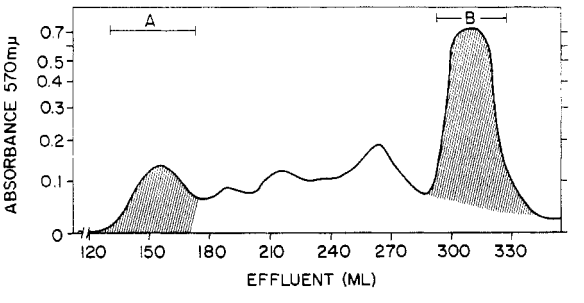


FIGURE 2: Elution profile of CNBr fragments of Hp β chain separated on Sephadex G-100 (2 × 100 cm column) in 30% acetic acid. The effluent was monitored on an automatic peptide analyzer by ninhydrin reaction after alkaline hydrolysis. (A) CNBr fragment II. (B) The C- and N-terminal CNBr fragments plus reagent.

Carboxyl-Terminal Peptide Sequence Analysis. The octapeptide sequence of the C-terminal tryptic and chymotryptic peptides has been previously reported (Barnett *et al.*, 1970). The first four residues of the nine-residue, C-terminal CNBr fragment were determined on a Beckman sequencer (Edman and Begg, 1967) by the 4-SPITC method (Braunitzer *et al.*, 1970) described for the N-terminal tryptic peptide. The 4-SPTH derivatives of the first and fourth residues were identified as their parent amino acid by amino acid analysis after alkaline hydrolysis. The second and third residues were identified as PTH derivatives by gas chromatography.

Automated Sequence Determination of the N-Terminal CNBr Fragment. The sequence of the first 8 residues of the 18 residue N-terminal CNBr fragment was determined on the Beckman sequencer using the DMAA PEPTIDE program 080570. Identification of these eight residues was by gas chromatography. The remaining decapeptide was completely washed out of the sequencer cup after the eighth degradation cycle.

Amino-Terminal Peptide Sequence Analysis. The sequences of the N-terminal tryptic and chymotryptic peptides were determined by Edman degradation (Edman, 1950) as modified by Gray (1967). Sequentially exposed N-terminal residues were identified as DNS derivatives by the method of Morse and Horecker (1966) except that polyamide plates were used for thin-layer chromatography (Woods and Wang, 1967). Subtractive amino acid analysis was carried out after degradation steps. The results of these experiments are found in Table II.

Automated Sequence Determination of the N-Terminal Tryptic Peptide. The N-terminal tryptic nonapeptide was subjected to automated sequence analysis in a Beckman sequencer using the 4-SPITC method of Braunitzer *et al.* (1970). About 500 nmoles of peptide were dried in a small glass-stoppered conical tube. After the tube was flushed with nitrogen the peptide was dissolved in 15 μl of water. To this was added 1 mg of 4-SPITC (Pierce Chemical Co.) and 30 μl of dimethylaminopropyne (Aldrich Chemical Co., Inc.). The reaction was carried out at 40° for 1 hr followed by addition of 5 μl of tributylamine (Aldrich Chemical Co., Inc.). The reaction was then allowed to proceed for an additional 1 hr at 40°. The tube was flushed with nitrogen after each addition of reagent. The peptide reaction mixture was diluted with 500 μl of water, introduced into the sequencer reaction cup, and dried in the usual manner. The dried film was extracted with butyl chloride for 300 sec by manual manipulation of the sequencer controls. In the first cycle, a second addition of PITC was made prior to introduction of quadrol.

The program used for the sequencer was Beckman QUADROL DOUBLE-CLEAVAGE "C," with the following modifications. (a) Quadrol buffer was made 0.5 M by a 1:1 dilution with water. (b) Ethyl acetate extraction was reduced from 600 to 300 sec. (c) The second cleavage and extraction steps were eliminated. (d) Butyl chloride contained 1 μl/ml of ethanethiol (Eastman Organic Chemicals). In conversion of thiozolinone amino acid derivatives to PTH derivatives, 1 μl/ml of ethanethiol was included in the 1 M HCl and in the ethyl acetate extraction solvent. Ethanethiol in the same concentration was included in the ethyl acetate used for injection into the gas chromatographic column.

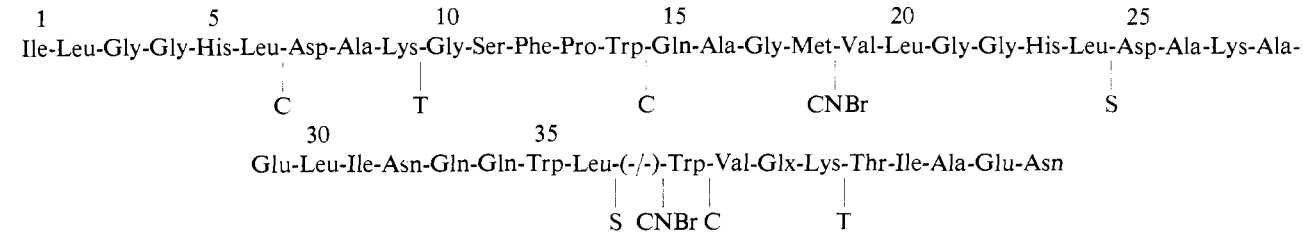
Identification of the PTH derivatives was made by gas chromatography according to the single-column method of Pisano and Bronzert (1969, 1970) using a Nuclear-Chicago, series 5000, gas chromatograph. Some amino acid residues were verified by amino acid analysis after hydrolysis to parent amino acids (Africa and Carpenter, 1966; O. Smithies, D. M. Gibson, and E. M. Fanning, 1971, personal communication). The cited methods for hydrolysis were modified as follows. (1) The 0.2 M NaOH contained 1 μl/ml of ethanethiol. (2) Hydrolysis was carried out at 130° for 2.5 hr in individual hydrolysis tubes equipped with a Teflon valve seal and with a side arm for drawing a vacuum and flushing with nitrogen. The tubes were specially fabricated by Glenco Scientific, Inc., Houston, Texas.

TABLE II: Stepwise Edman Degradation of the N-Terminal Tryptic (T-5-E) and Chymotryptic (C-14-B) Peptides and the Second Chymotryptic Peptide (C-9-B).

Sequence	Tryptic Peptide T-5-E Ile-Leu-Gly-Gly-His-Leu-Asx-Ala-Lys
Edman degradation[a]	
Step 0 (intact peptide)	DNS-Ile
Step 1	DNS-Leu
Step 2	DNS-Gly
Step 3	DNS-Gly; residue: Lys, 0.7 (1); His, 0.9 (1); Asp, 1.0 (1); *Gly, 1.1 (1)*; Ala, 0.9 (1); *Ile, 0; Leu, 0.9 (1).*
Step 4	No DNS-amino acid identified. His by subtraction.
Step 5	DNS-Leu; residue: Lys, 0.6 (1); *His, 0.2 (0)*; Asp, 1.0 (1); *Gly, 0.3 (0)*; Ala, 0.9 (1); Leu, 0.9 (1.0).
Step 6	DNS-Asp (Asx)
Step 7	DNS-Ala
Step 8	Lys by subtraction.
	Chymotryptic Peptide C-14-B Ile-Leu-Gly-Gly-His-Leu
Step 0	DNS-Ile
Step 1	DNS-Leu; residue: His, 0.9 (1); Gly, 1.9 (2); *Ile, trace (0)*; Leu, 2.0 (2).
Step 2	DNS-Gly
Step 3	DNS-Gly; residue: His, 1.1 (1); Gly, 1.2 (1); *Leu*, 1.0 (1).
Step 4	No DNS-amino acid identified. His by subtraction.
Step 5	DNS-Leu; residue: *His, 0; Gly, trace (0)*; Leu, 1.0 (1).
	Chymotryptic Peptide C-9-B Asx-Ala-Lys-Gly-Ser-Phe-Pro-Trp
Step 0	DNS-Asp (Asx)
Step 1	DNS-Ala
Step 2	No DNS-amino acid identified. Lys by subtraction.
Step 3	DNS-Gly
Step 4	DNS-Ser; Residue: *Lys, 0; Asp, 0;* Ser, 0.8 (1); Pro, 0.8 (1); *Gly, 0.1 (0)*; *Ala, 0;* Phe, 1.0 (1).
Step 5	DNS-Phe
Step 6	DNS-Pro
Step 7	Trp by subtraction.

[a] N-Terminal residues identified as DNS derivatives. Residues missing by subtractive analysis are in italic type.

TABLE III. Amino Acid Sequence of the N- and C-Terminal Regions of the Haptoglobin β Chain.[a]



[a] Points of various cleavage reactions are indicated: C, chymotrypsin; T, trypsin; CNBr, cyanogen bromide; S, sequencer (Residues 1–24 and 19–36). -/- indicates residues between the N- and C-terminal sequences.

Automated Sequence Determination of the N-Terminal Region of the Intact Hp β Chain. The intact Hp β chain was subected to automated sequence analysis (Edman and Begg, 1967) on a Beckman sequencer using the Beckman QUADROL DOUBLE CLEAVAGE program "C". Ethanethiol (1 μl/ml) was included in the butyl chloride extraction solvent. Conversion of thiozolinone amino acid derivatives from the sequencer was carried out as described above. The resulting PTH derivatives were identified by gas chromatographic analysis and amino acid analysis.

Automated Sequence Determination of CNBr Fragment II. The sequence of the first 18 residues of CNBr fragment II was determined on the automatic sequencer.

Results

Haptoglobin Isolation and β-Chain Purification. The α and β chains from the three haptoglobin phenotypes were completely separated on a guanidine Sephadex G-200 column (Gordon *et al.*, 1968) (Figure 1). The sequence analysis of β-chain preparations indicated identical sequences regardless of the haptoglobin phenotype (Hp 1-1, 2-1, or 2-2) or source of the haptoglobin (plasma or ascites fluid). The sequences of the N- and C-terminal tryptic peptides were determined on preparations obtained from plasma of Hp 2-1. The chymotryptic peptides were isolated from a preparation obtained from plasma of Hp type 1-1. The N-terminal sequence of the intact Hp β chain was determined on preparations from ascites fluid of Hp type 1-1 and 2-2. The CNBr fragments used in this study were obtained from preparations of ascites fluid of Hp 1-1 and 2-2.

Carboxyl-Terminal Labeling. Amino acid analysis of a β-chain preparation which had been reacted with [^{3}H]water and digested with carboxypeptidase followed by acid hydrolysis revealed that 78% of the radioactivity was incorporated in the C-terminal residue, asparagine. The peak characteristic for asparagine, glutamine, and serine, which was radioactive, shifted to that characteristic for aspartic acid after acid hydrolysis. The amino acid composition of this β-chain preparation is shown in Table I. The number of residues were calculated as described by Cleve *et al.* (1969); the results closely agree with those reported. After the same ^{3}H-labeled preparation was reacted with CNBr the C-terminal CNBr nonapeptide was isolated as described.

Cyanogen Bromide Fragmentation. Under the conditions described above the cleavage by CNBr appeared to be complete. No fragment was obtained which demonstrated the presence of methionine or more than one residue of homoserine. Furthermore, six fragments were isolated as would be expected from the complete cleavage of a polypeptide containing five residues of methionine. The detailed description of the CNBr fragments will be presented in a subsequent paper. The separated fragments obtained by gel filtration on Sephadex G-100 are shown in Figure 2. Those fragments used in this study are indicated by shaded areas of the elution profile and include fragment I, the N-terminal fragment; fragment II, which is adjacent to the N-terminal in the polypeptide chain; and the C-terminal fragment. The amino acid composition of these fragments is shown in Table I.

Sequence of the C-Terminal CNBr Fragment. The sequence of the first four residues of the nine-residue, C-terminal CNBr fragment was found to be Trp-Val-Glu-Lys. The detailed analysis of the overlapping tryptic hexapeptide and chymotryptic octapeptide has been reported elsewhere (Barnett *et al.*, 1970). The complete sequence of these overlapping peptides is Trp-Val-Glu-Lys-Thr-Ile-Ala-Glu-Asn.

Partial Sequence of the N-Terminal CNBr Fragment. Automated sequence analysis using the Beckman DMAA PEPTIDE program and gas chromatographic analysis of the N-terminal CNBr fragment revealed the first eight residues to be Ile-Leu-Gly-Gly-His-Leu-Asp-Ala. After the eighth cycle of degradation the remaining peptide material was completely washed out of the reaction cup in extraction solvents.

Edman Degradation of the N-Terminal Tryptic and Chymotryptic Peptides. The amino acid composition of the N-terminal tryptic (T-5-E) and overlapping chymotryptic peptides (C-14-B; C-9-B) is shown in Table I. The details of the analysis of the DNS derivatives and subtractive Edman degradation data are shown in Table II.

Automated Sequence Determination of the N-Terminal Tryptic Peptide. Use of 4-SPITC made it possible to determine the complete sequence of the N-terminal tryptic peptide on an automatic sequencer. The N-terminal residue, isoleucine, was

TABLE IV: Comparison of the C-Terminal Sequences of Hp β Chain, Bovine Chymotrypsin A (BC-A), and Bovine Trypsin (BT).[a]

Hp β	*Trp-Val-Glx*-Lys-*Thr-Ile-Ala*-Glu-*Asn*
BCA	*Trp-Val-Gln*-Gln-*Thr*-Leu-*Ala*-Ala-*Asn*
BT	*Trp*-Ile-Lys-Gln-*Thr-Ile-Ala*-Ser-*Asn*

[a] Identical residues are in italic type.

TABLE V: Comparison of the N-Terminal Sequences of Hp β Chain, Bovine Chymotrypsin-A (BC-A), and Bovine Trypsin (BT).[a]

	1 5 10 11 15 18
Hp β	*Ile*-Leu-*Gly*-*Gly*-His-Leu-Asp-Ala-Lys-*Gly*-*Ser*-Phe-*Pro*-*Trp*-*Gln*-Ala-Gly-Met
BC-A	*Ile*-Val-Asn-*Gly*-Glu-Glu-Ala-Val-Pro-*Gly*-*Ser*-Trp-*Pro*-*Trp*-*Gln*-Val-Ser-Leu
BT	*Ile*-Val-*Gly*-*Gly*-Tyr-Thr-Cys-Gly-Ala-Asn-Thr-Val-*Pro*-Tyr-*Gln*-Val-Ser-Leu

[a] Identical residues are in italic type.

identified as isoleucine and *allo*-isoleucine by amino acid analysis. Analysis of the PTH derivatives by gas chromatography revealed the presence of leucine, glycine, and glycine in the second, third, and fourth positions, respectively. In the fifth position histidine was identified by amino acid analysis and by specific staining. Leucine was identified as its PTH derivative by gas chromatography. The trimethylsilylphenyl isothiocyanate derivative of aspartic acid was identified by gas chromatography. After degradation cycle seven the remaining dipeptide was shown to contain alanine and lysine by amino acid analysis. Lysine was assumed to be C-terminal due to trypsin specificity. The sequence of the N-terminal tryptic peptide as determined on an automatic sequencer is Ile-Leu-Gly-Gly-His-Leu-Asp-Ala-Lys.

N-Terminal Sequence of the Intact Hp β Chain. The first 24 residues of the intact Hp β chain (Table III) were obtained by analysis of PTH derivatives from the automatic sequencer. This sequence provided an overlap between the N-terminal CNBr fragment, fragment I, and CNBr fragment II. The overlap is observed in residues 19–24.

Partial Sequence of CNBr Fragment II. Residues 19–36 of the Hp β chain were determined by automated analysis of CNBr fragment II and are shown in Table III.

Discussion

The presence of methionine residues, 1 within the first 18 amino acid residues of the N-terminal region of the Hp β chain and the other preceding the 9 C-terminal residues permitted the separation of the N- and C-terminal fragments from the remaining CNBr fragments. The sequence of these two regions of the Hp β chain has been established utilizing the sequences of tryptic peptides, chymotryptic peptides, cyanogen bromide fragments, the amino-terminal sequence of the intact β chain, and the carboxyl-terminal sequence of the ^{3}H-labeled C-terminal CNBr fragment.

When these sequences of the Hp β chain were compared to published N- and C-terminal sequences of bovine chymotrypsin A (Hartley and Kauffman, 1966) and bovine trypsin (Walsh *et al.*, 1964), some unexpected homologies were obvious. In the carboxyl-terminal sequences of the haptoglobin β chain and bovine chymotrypsin A six of the nine residues are identical (Table IV). The three nonidentical residues could be explained as a result of a single nucleotide base change in the codon. When the β-chain sequence is compared to that of bovine trypsin, five of the nine residues are identical. Of the remaining four, three could be explained by a single nucleotide base change in the codon, while the alteration of glutamic acid to serine, in the penultimate position, would require two base changes in the codon.

When the amino-terminal sequence of the first 18 residues of Hp β chain, bovine chymotrypsin A, and bovine trypsin are compared, homology is again obvious (Table V). Of the 18 residues there are 7 identical positions between the β chain and bovine chymotrypsin A, while there are 5 identical residues between the β chain and bovine trypsin. The homology demonstrated between these sequences of the Hp β chain and bovine chymotrypsin A is only slightly less than that observed between bovine chymotrypsin A and bovine trypsin. The observed homology between Hp β chain and the two proteinases was unexpected since haptoglobin is a glycoprotein without demonstrated proteolytic activity. Furthermore, the haptoglobin quaternary structure is comprised of two pairs of dissimilar chains connected through disulfide bonds. Bovine chymotrypsin A and bovine trypsin, on the other hand, are single polypeptide chains which do not contain carbohydrate. However, the β chain of bovine thrombin, a proteolytic serum protein known to have a primary sequence homologous with chymotrypsin and trypsin, is comprised of subunits and has a carbohydrate moiety attached to its heavy chain. Thrombin has an N-terminal sequence similar to those proteins shown in Table V, *i.e.*, Ile-Val-Glu-Gly-Gln-Asp (Magnusson, 1970). One interesting and unexplained feature of the observed homologies between the Hp β chain, chymotrypsin, and trypsin is that the sequence similarities occur in the N-terminal regions of the proteolytic enzymes and not in the N-terminal regions of the zymogens. This is consistent with the observed homologies within the serine proteinases and lack of homologies in the N-terminal regions of their zymogens.

In the sequence determination of the smaller haptoglobin subunit, the α chain, Black and Dixon (1968) found sequence similarities to the Bence-Jones κ and λ chains of the immunoglobulins. Therefore, haptoglobin demonstrates sequence homology between its α chains and immunoglobulins in addition to homology between the β chains and the serine proteinases. The observed similarities may reflect that the structural genes directing the synthesis of the Hp α and β chains were each derived from ancestral genes common to those for the immunoglobulins and serine proteinases, respectively. An alternate explanation, although less likely, is that the genes coding for the α and β chains have arisen independently from those of the immunoglobulins and proteinases and due to random events, evolved convergently.

The N-terminal sequence of the Hp β chain demonstrated an internal homology within the first 27 residues (Table VI). The sequence of residues 2–9 is identical with the sequence of residues 20–27, while residue 1, isoleucine, is similar to residue 19, which is valine. It is interesting that the repeated sequence involves the N-terminal sequence found to be homologous with the serine proteinases. Because of this it is tempting to suggest that the genetic event leading to the internal homology involved a nonhomologous crossover between genes coding for the serine proteinases. Alternately, a partial gene duplication of the Hp$^\beta$ gene may have occurred after the divergence of genes coding for the Hp β chain and the serine proteinases.

In haptoglobin an evolutionary event involving partial gene

TABLE VI: Illustration of Internal Homology within the Haptoglobin β Chain.[a]

1			5			9		
Ile-	*Leu-*		*Gly-*	*Gly-*	*His-*	*Leu-*	*Asp-*	*Ala-Lys*
19					25		27	
Val-	*Leu-*		*Gly-*	*Gly-*	*His-*	*Leu-*	*Asp-*	*Ala-Lys*

[a] Identical residues are in italic type. Numbers represent the position of residues in the intact Hp β chain.

duplication has been proposed (Smithies *et al.*, 1962a) and substantiated (Black and Dixon, 1968) by direct experimental evidence, for the production of the Hp^{α} allele, Hp^{α_2}, which codes for a polypeptide almost double the length of the gene product of Hp^{α_1}. The work of Black and Dixon (1968) has demonstrated that a partial gene duplication had occurred such that the original gene product, an α chain of 83 residues, was altered to a polypeptide of 142 residues in which the sequence was almost completely repeated.

The complete sequence of the Hp β chain will clarify the genetic mechanisms leading to the evolution of the Hp^{β} gene and its relationship to other classes of proteins.

Acknowledgments

We are grateful to Mr. Horace D. Kelso for valuable technical assistance and to Dr. Marilyn Rasco for assistance in the ^{3}H-labeling experiments. We thank Dr. Alvin LeBlanc, University of Texas Medical Branch, and Dr. Julian Smith, M. D. Anderson Hospital, for ascites fluid. The preparation of 4-SPITC was synthesized by Pierce Chemical Co. and was a generous gift from Mr. Charles Lantz.

References

Africa, B., and Carpenter, F. H. (1966), *Biochem. Biophys. Res. Commun. 24*, 113.

Barnett, D. R., Lee, T.-H., and Bowman, B. H. (1970), *Nature (London) 225*, 938.

Black, J. A., and Dixon, G. H. (1968), *Nature (London) 218*, 736.

Braunitzer, G., Schrank, B., and Ruhfus, A. (1970), *Hoppe-Seyler's Z. Physiol. Chem. 351*, 1589.

Cleve, H., Bowman, B. H., and Gordon, S. (1969), *Human. genetik 7*, 337.

Cleve, H., Gordon, S., Bowman, B. H., and Bearn, A. G. (1967), *Amer. J. Hum. Gen. 19*, 713.

Edman, P. (1950), *Acta Chem. Scand. 4*, 283.

Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1*, 80.

Gordon, S., and Bearn, A. G. (1966), *Proc. Soc. Exp. Biol. Med. 121*, 846.

Gordon, S., Cleve, H., and Bearn, A. G. (1968), *Proc. Soc. Exp. Biol. Med. 127*, 52.

Gray, W. R. (1967), *Methods Enzymol. 11*, 469.

Gross, E. (1967), *Methods Enzymol. 11*, 238.

Hartley, B. S., and Kauffman, D. L. (1966), *Biochem. J. 101*, 229.

Holcomb, G. N., James, S. A., and Ward, D. N. (1968), *Biochemistry 7*, 1291.

Jayle, M. F., and Moretti, J. (1962), *Progr. Hematol. 3*, 342.

Jones, R. T. (1966), Presented at the Technicon Symposium "Automation in Analytical Chemistry", New York, N.Y., October 18, 1966.

Jordan, E. M., and Raymond, S. (1969), *Anal. Biochem. 27*, 205.

Magnusson, S. (1970), *in* Structure-Function Relationships of Proteolytic Enzymes, Desnuelle, P., Neurath, H., and Ottesen, M., Ed., Copenhagen, Munksgaard, p 138.

Matsuo, H., Fujimoto, Y., and Tatsuno, T. (1966), *Biochem. Biophys. Res. Commun. 22*, 69.

Morse, D., and Horecker, B. L. (1966), *Anal. Biochem. 14*, 429.

Nagel, R. L. (1967), *J. Biol. Chem. 242*, 3428.

Pisano, J. J., and Bronzert, T. J. (1969), *J. Biol. Chem. 244*, 5597.

Pisano, J. J., and Bronzert, T. J. (1970), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 29*, 916.

Polonovski, M., and Jayle, M. F. (1938), *C. R. Soc. Biol. 129*, 457.

Shim, B.-S., and Bearn, A. G. (1964), *J. Exp. Med. 120*, 611.

Shim, B.-S., Lee, T.-H., and Kang, Y.-S. (1965), *Nature (London) 207*, 1264.

Smithies, O., Connell, G. E., and Dixon, G. H. (1962a), *Nature (London) 196*, 232.

Smithies, O., Connell, G. E., and Dixon, G. H. (1962b), *Amer. J. Hum. Genet. 14*, 14.

Wada, T., Ohara, H., Watanabe, K., Kinoshita, H., and Nishio, H. (1970), *J. Retic. Soc. 8*, 195.

Walsh, K. A., Kauffman, D. L., Kumar, K. S. V. S., and Neurath, H. (1964), *Proc. Nat. Acad. Sci. U. S. 51*, 301.

Woods, K. R., and Wang, K.-T. (1967), *Biochim. Biophys. Acta 133*, 369.