Complete Amino Acid Sequence of a Papain-Solubilized Human Histocompatibility Antigen, HLA-B7. 1. Isolation and Amino Acid Composition of Fragments and of Tryptic and Chymotryptic Peptides†

José A. López de Castro, Harry T. Orr, Richard J. Robb, Thomas G. Kostyk, Dean L. Mann, and Jack L. Strominger*

ABSTRACT: As a part of the overall strategy for determining the complete covalent structure of the papain-solubilized portion of the heavy chain of the human histocompatibility antigen HLA-B7, the protein was dissected into various fragments by a combination of partial acid hydrolysis and cyanogen bromide cleavage. After purification by chromatographic procedures, these fragments have been used as a source for tryptic and chymotryptic peptides. Thirty-three major tryptic and twenty-two major chymotryptic peptides were purified in nanomole amounts and their amino acid compositions determined. These peptides account for the whole extent of the polypeptide chain with the exception of the amino-terminal CNBr pentapeptide. They provide the basis for the formal alignment of the acid cleavage and cyanogen bromide fragments of the molecule as well as the source material for the elucidation of the primary structure of the HLA-B7 heavy chain.

The HLA major histocompatibility complex (MHC)[1] is a genetic region located in human chromosome 6 (van Someran et al., 1974). These genes control the rejection of tissue allografts and appear to play a central role in controlling a number of immune functions (Paul & Benacerraf, 1977; Thorsby, 1978). The HLA-A, -B, and -C loci control the expression of ubiquitous cell surface glycoproteins known as histocompatibility antigens because of their role as the major antigenic structures responsible for graft rejection. In addition, they are involved in the recognition of targets by cytotoxic T lymphocytes (Dickmeiss et al., 1977; Tursz et al., 1977; McMichael et al., 1977).

Histocompatibility antigens display an extraordinary genetic polymorphism which is probably of great functional significance. Each one of the HLA-A, -B, and -C loci contains a large number of alleles which are codominantly expressed. These alloantigens are integral membrane proteins composed of two polypeptide chains held together by noncovalent interactions (Cresswell et al., 1973). The light chain (~12000 daltons) is not encoded by the MHC, is not polymorphic, and has been shown to be identical with β_2-microglobulin (Poulik & Reisfeld, 1976). The heavy chain (~44000 daltons) is encoded by the MHC and constitutes the polymorphic gene product of the HLA-A, -B, and -C loci.

Treatment of cell membranes with papain results in the solubilization of an antigenically active portion of the histocompatibility antigen which consists of ~80% of the native molecule, including β_2-microglobulin and the amino-terminal 75% of the heavy chain (Sanderson, 1968; Cresswell et al., 1974b; Springer & Strominger, 1976). The complex behaves as a stable globular protein, readily soluble in aqueous solutions, which may be purified in the absence of detergents by conventional chromatographic techniques (Parham et al., 1977).

In the present communication, the strategy for large-scale purification of polypeptide fragments from the papain-soluble portion of HLA-B7 heavy chain is described, as well as the characterization of tryptic and chymotryptic peptides. A definitive alignment of the fragments is provided, confirming previous reports (Terhorst et al., 1977; Trägårdh et al., 1979). The following paper (Orr et al., 1979) presents the complete sequence of the papain-solubilized HLA-B7 heavy chain.

Materials and Methods

Preparation of Papain-Solubilized HLA-B7 Heavy Chain. Cells of the human lymphoblastoid cell line JY were used as a source of material (Terhorst et al., 1976). The cells have been repeatedly typed as homozygous for HLA-A2 and HLA-B7. Preparation of membranes, papain solubilization, and purification of HLA-B7 antigen have been described in a previous report (Parham et al., 1977). Briefly, membranes were obtained after hypotonic lysis of cells in 0.01 Tris-HCl, pH 8.0, washed in the same solution, and treated with papain (Worthington, 6000 units of enzyme per 100 g of cells). The enzyme was subsequently removed by DEAE-cellulose chromatography. The solubilized material was fractionated by gel filtration using a Sephacryl S-200 (Pharmacia) column equilibrated with 0.14 M NH_4HCO_3. Fractions containing HLA-A and -B antigens were pooled and subsequently fractionated by DEAE-cellulose chromatography using a linear gradient from 0.005 M Tris–phosphate, pH 8.0, to 0.1 M Tris–phosphate, pH 5.6, to separate HLA-2 and HLA-B7 specificities. HLA-B7 antigen preparations were judged to be homogeneous by polyacrylamide gel electrophoresis in sodium dodecyl sulfate.

Complete reduction and [^{3}H]carboxymethylation were carried out essentially as described (Parham et al., 1977). Separation of HLA-B7 heavy chain and β_2-microglobulin was done by gel filtration in a column (1.6 × 100 cm) of Bio-Gel P-300 in 6 M guanidine hydrochloride. Fractions containing

† From the Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138 (J.A.L.C., H.T.O., R.J.R., T.G.K., and J.L.S.), and the National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205 (D.L.M.). *Received July 17, 1979.* Supported by a research grant from the National Institutes of Health (AI 10736). H.T.O. has a U.S. Public Health Service Research Fellowship (AI 05483), and J.A.L.C. has a U.S. Public Health Service International Research Fellowship (TW 02582).

[1] Abbreviations used: MHC, major histocompatibility complex; $NaDodSO_4$, sodium dodecyl sulfate; CNBr, cyanogen bromide; DEAE, diethylaminoethyl; TPCK, 1-(tosylamido-2-phenyl)ethyl chloromethyl ketone.

0006-2960/79/0418-5704$01.00/0

the heavy chain were pooled, dialyzed against 0.1 M NH_4HCO_3, and lyophilized. Alternatively, the reduced and alkylated protein was applied directly to a column (1.6 × 100 cm) of Sephadex G-75 equilibrated with 1 M acetic acid. Fractions containing the heavy chain were pooled and lyophilized. HLA-B7 heavy-chain preparations were homogeneous by $NaDodSO_4$ gel electrophoresis and subsequent sequence analysis.

Preparation of Fragments by Mild Acid Hydrolysis. About 5 mg of reduced and [^{3}H]carboxymethylated protein was dissolved in 2 mL of 70% formic acid containing 0.01% (v/v) β-mercaptoethanol, and nitrogen was bubbled through the solution. These precautions proved to be necessary in order to prevent oxidation of methionine residues. The mixture was incubated for 36 h at 37 °C. At the end of this period, the sample was applied directly to a Sephadex G-75 column (1.6 × 100 cm) equilibrated with 1 M acetic acid. Fractions containing the purified fragments were pooled and lyophilized.

Preparation of CNBr Fragments. Lyophilized, reduced, and [^{3}H]carboxymethylated HLA-B7 heavy chain was cleaved by CNBr as described (Gross, 1967). After 24 h, the reaction mixture was diluted in 10 volumes of water and lyophilized. This material was dissolved in 2 mL of 0.2 M NH_4HCO_3, pH 8.5, and applied to a column (0.9 × 30 cm) of DEAE-cellulose (Whatman DE-52) equilibrated with the same buffer. The column was first eluted with this buffer until A_{280} reached background (~3 column volumes). At this point, elution was continued with 0.3 M NH_4HCO_3, pH 8.5, and then successively with 0.4 and with 0.5 M NH_4HCO_3, pH 8.5. Fractions were monitored by reading absorbance at 280 nm. Distribution of radioactivity was detected by liquid scintillation counting aliquots of each fraction in Aquasol (New England Nuclear). Protein fractions recovered at each elution step were pooled and further fractionated by gel filtration on a Sephadex G-75 column (1.6 × 100 cm) equilibrated with 1 M acetic acid. Fractions were monitored as above.

In some preparations, protein fragments obtained after mild acid hydrolysis were further cleaved with CNBr, following the same procedure (Gross, 1967). These fragments were fractionated by a single gel filtration step using a Sephadex G-75 column (1.6 × 100 cm) equilibrated with 1 M acetic acid.

Polyacrylamide Gel Electrophoresis. Sodium dodecyl sulfate ($NaDodSO_4$)–polyacrylamide gel electrophoresis, as described by Laemmli (1970), was routinely used to monitor the purification of HLA-B7 antigen. The method described by Maizel (1971) was used for the analysis of protein fragments. Gels were stained with Coomassie Brilliant Blue and destained by repeated washing with 7.5% acetic acid and 10% methanol in water.

Tryptic and Chymotryptic Digestion. Protein samples were dissolved (1–5 mg/mL) in 0.05 M NH_4HCO_3, pH 8.5. TPCK-treated trypsin (Worthington) was added in 0.001 N HCl at an enzyme/substrate ratio of 1:100 3 times at 2-h intervals. Incubation was carried out at 37 °C for a total period of 6 h. After digestion, the sample was immediately lyophilized.

Chymotryptic digestion was carried out by dissolving protein samples as above and subsequently adding α-chymotrypsin (Worthington) in 0.001 N HCl at an enzyme/substrate ratio of 1:200. The digestion mixture was incubated at room temperature for 30 min and then immediately lyophilized.

Peptide Purification. Microbore ion-exchange chromatography (Machleidt et al., 1977) was used as the main procedure for purification of preparative amounts of peptides throughout this study. A Beckman column (0.28 × 20 cm) equipped with a water jacket was packed with 1 to 2 g of Beckman AA20 spherical resin under nitrogen pressure. The column was packed and maintained at 52 °C by using a circulator water bath (Haake, Model FE). After being packed, the resin was conditioned by successive washes of several column volumes with 3 M NaOH, H_2O, 3 M HCl, H_2O, and 8 M pyridine (Herman & Vanaman, 1977). Finally, the column was equilibrated with 0.05 M pyridine–acetic acid buffer, pH 1.8.

Routinely, 50–100 nmol of digested material was dissolved in 0.5 mL of the equilibrating buffer and applied to the column. The column was then eluted with the same buffer, and 20 fractions of 1 mL each were collected. Elution was then continued with a gradient formed in a four-chamber mixer containing 50 mL/chamber of the following buffers: chamber 1, 0.05 M pyridine–acetic acid, pH 2.5; chambers 2 and 3, 0.3 M pyridine–acetic acid, pH 3.8; chamber 4, 1.2 M pyridine–acetic acid, pH 5.0. After the gradient, a further elution with 2 M pyridine was carried out for 20 additional fractions. The column was then washed with 20 column volumes of 8 M pyridine and reequilibrated with starting buffer.

Peptide detection in the effluent fractions was performed by using the fluorescamine assay as described by Lai (1977). Because of the sensitivity of this method and the small amounts of peptides being separated, both reagents used in the chromatographic buffers were redistilled. Acetic acid (Fisher) was refluxed over Cr_2O_3 and then distilled once. Pyridine (Fisher) was refluxed over nitrogen and then distilled 3 times and stored at 4 °C under nitrogen. This procedure consistently yielded low fluorescent backgrounds. ([^{3}H]Carboxymethyl)cysteine-containing peptides were detected by liquid scintillation counting of 20-μL aliquots of each fraction in Aquasol.

Occasionally, simple mixtures of peptides which coeluted by microbore ion-exchange chromatography were resolved by preparative thin-layer chromatography (Gracy, 1977). Peptides were visualized by fluorescamine staining and extracted with 0.1 M acetic acid.

The purification of a tryptic glycopeptide obtained after digestion of the whole polypeptide chain which provides the overlap for two CNBr fragments has been described elsewhere (Parham et al., 1977).

Amino Acid Analysis. Peptide samples were hydrolyzed at 110 °C for 24 h in constant boiling 5.7 N HCl (Pierce) containing 1 μg/mL 2-mercaptoethanol. When quantitation of methionine was required, phenol (1% v/v) was added instead of 2-mercaptoethanol. Amino acid analysis was performed on a Beckman 121M analyzer.

Results

Chemical Cleavage of Papain-Solubilized Heavy Chain. The strategy used in the sequence determination of HLA-B7 heavy chain was based on the isolation of large fragments derived by specific chemical cleavage. A combination of selective acid hydrolysis at the single aspartyl–proline bond of the molecule (Terhorst et al., 1977) and CNBr cleavage was used. A general scheme of this purification is given in Figure 1. The separation of fragments obtained after partial acid hydrolysis of the polypeptide chain is illustrated in Figure 2A. Under conditions of this reaction (see Materials and Methods) as much as 90% of the protein is specifically cleaved at the susceptible aspartyl–proline bond, with no significant amount of breakdown byproducts arising from nonspecific acid cleavage at other peptide bonds.

ac-1 fragment (peak 2 in Figure 2A) was further cleaved by CNBr. The yield of this reaction was as high as 80% as long as care was taken to avoid oxidizing conditions during

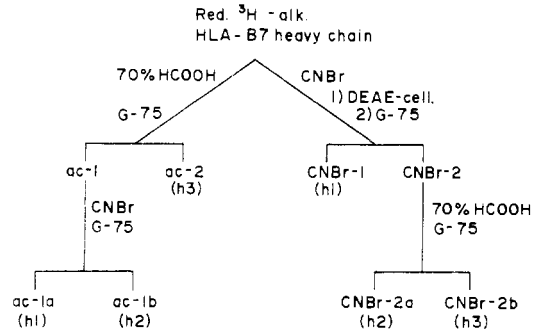

FIGURE 1: Flow chart showing the purification scheme of acid cleavage and CNBr fragments from papain-solubilized HLA-B7 heavy chain. The nomenclature reflects the primary cleavage method used in the generation of the fragments. Thus, ac reflects acid cleavage and CNBr means cyanogen bromide cleavage. A final designation for equivalent fragments is given in parentheses.

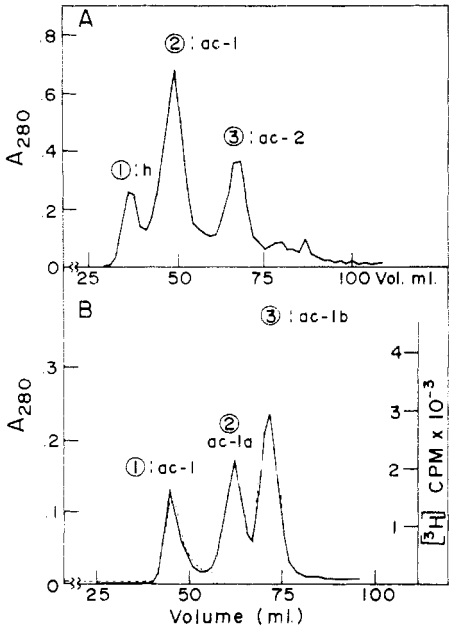

FIGURE 2: (A) Purification of fragments obtained after partial acid hydrolysis of papain-solubilized, reduced, and [^{3}H]carboxymethylated HLA-B7 heavy chain. A G-75 Sephadex column (1.6 × 100 cm) equilibrated with 1 M acetic acid was used at a flow rate of 3 mL/h. Fractions of 1.5 mL were collected and monitored by reading absorbance at 280 nm (A_{280}). (B) ac-1 fragments (peak 2 in Figure 2A) were cleaved by CNBr and fractionated by gel filtration as above. Fractions were monitored for A_{280} (—) and for ([^{3}H]carboxymethyl)cysteine radioactivity (---). Uncleaved material (ac-1) elutes as peak 1, and two other peaks are obtained: ac-1a (peak 2) contains the carbohydrate moiety of the molecule; ac-1b (peak 3) contains all the ([^{3}H]carboxymethyl)cysteine present in the parental fragment ac-1.

the previous acid cleavage reaction (see Materials and Methods). The separation of the resulting fragments is shown in Figure 2B. Peak 1 of this chromatography represents uncleaved material (ac-1). Peak 2 (ac-1a) contains carbohydrate as judged by amino acid analysis and contains no radioactivity. Peak 3 (ac-1b) contains 2 residues/mol of ([^{3}H]carboxymethyl)cysteine and does not contain carbohydrate.

A second strategy of fragmentation was developed for papain-solubilized HLA-B7 heavy chain. After reduction and [^{3}H]carboxymethylation, the protein was subjected to CNBr cleavage. The reaction was performed at room temperature for 24 h. Typically, 80% of the protein was cleaved. As much as 10–15% acid cleavage at the susceptible aspartyl–proline bond occurred under these conditions. The reaction mixture was totally soluble in ammonium bicarbonate buffers and was

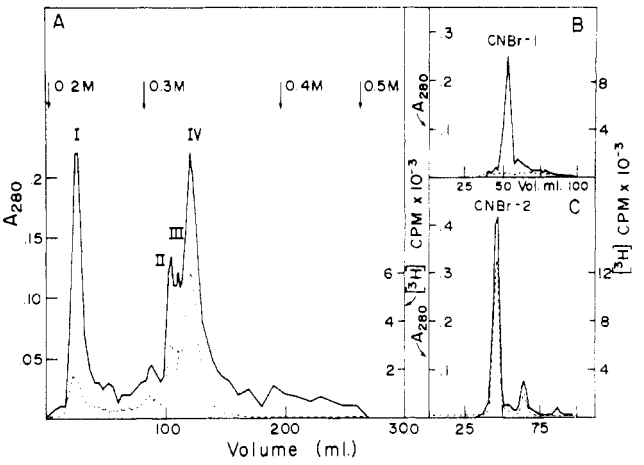

FIGURE 3: Purification of CNBr fragments. (A) Papain-solubilized HLA-B7 heavy chain was cleaved by CNBr and the resulting mixture fractionated by ion-exchange chromatography using a column (1.0 × 30 cm) of DEAE-cellulose. The column was equilibrated with 0.2 M NH_4HCO_3, pH 8.5. Elution was carried out initially with this buffer, and 1-mL fractions were collected. Fractions were monitored by measuring A_{280} (—) and ([^{3}H]carboxymethyl)cysteine radioactivity (---). After peak I was recovered, the eluting buffer was switched to 0.3 M NH_4HCO_3. With this buffer, peaks II, III, and IV were isocratically resolved. Elution with bicarbonate buffers of higher concentration did not result in further recovery of protein. (B) Peak I (Figure 3A) was further purified by gel filtration using a column (1.5 × 100 cm) of Sephadex G-75 equilibrated with 1 M acetic acid to eliminate minor contaminants. Fractions of 1.5 mL were collected at a flow rate of 3 mL/h and monitored as above. (C) Peak IV (Figure 3A) was purified by gel filtration as above to eliminate contaminating acid cleavage byproducts obtained during the CNBr reaction.

therefore fractionated by DEAE-cellulose chromatography as indicated in Figure 3A.

The protein fraction eluted with 0.2 M ammonium bicarbonate (peak I) was further purified by gel filtration (Figure 3B). A protein peak was recovered essentially free of radioactivity and containing carbohydrate as judged by amino acid analysis. This fragment was homogeneous by $NaDodSO_4$–polyacrylamide gel electrophoresis and was shown to be identical to ac-1a (Figure 2B) on the basis of its electrophoretic mobility and by amino-terminal sequence analysis. It was designated as h1 (see Figure 5).

Elution with 0.3 M ammonium bicarbonate resulted in three protein peaks (II, III, and IV in Figure 3A). Peaks II and III contained a mixture of uncleaved material and byproducts of the secondary acid cleavage reaction, as judged by $NaDodSO_4$–polyacrylamide gel electrophoresis. Peak IV contained mainly a single fragment with no carbohydrate and with high radioactivity. This fragment was further purified by gel filtration to separate a small amount of contaminant acid-cleavage byproducts (Figure 3C). The purified fragment (CNBr-2) was homogeneous by $NaDodSO_4$–polyacrylamide gel electrophoresis. This fragment contains the susceptible aspartyl–proline bond and therefore was further cleaved by partial acid hydrolysis. The yield of the reaction was ~80%, and the two fragments generated were separated by gel filtration (Figure 4). Both of these were radioactive, and each one contained 2 residues/mol of (carboxymethyl)cysteine (Table I). CNBr-2a fragment (peak 3 in Figure 4) was identical with ac-1b (Figure 2B) on the basis of amino acid composition. This fragment is designated as h2. CNBr-2b fragment (peak 2 in Figure 4) was identical by the same criterion with ac-2 (Figure 2A) and is designated h3.

A schematic representation of the papain-solubilized HLA-B7 heavy chain showing the different cleavage fragments is given in Figure 5. The evidence for the proposed alignment

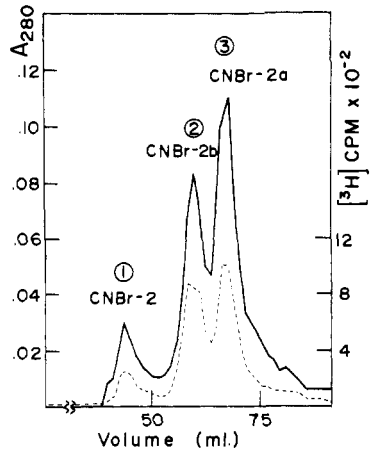

FIGURE 4: Fragment CNBr-2 (Figure 3C) was cleaved by partial acid hydrolysis. The resulting fragments (CNBr-2a and CNBr-2b) were resolved by gel filtration using a column (1.5 × 100 cm) of Sephadex G-75 equilibrated with 1 M acetic acid. Fractions of 1.5 mL were collected at a flow rate of 3 mL/h. A_{280} (–) and ([^{3}H]carboxymethyl)cysteine (---) were measured in monitoring the fractions.

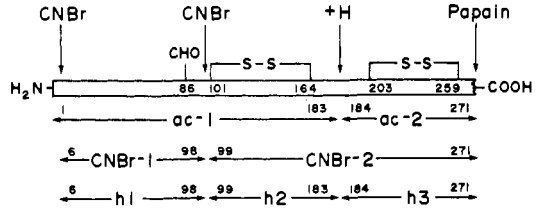

FIGURE 5: Schematic scale drawing of the papain-solubilized HLA-B7 heavy chain, showing the cleavage points by CNBr and by acid. The definitive alignment and dimensions of the fragments as derived by sequence analysis are given. The figures indicate residue numbers. The drawing shows the exact position of carbohydrate and the relative dimensions of the disulfide loops and the interloop region.

Table I: Amino Acid Composition of the Papain-Solubilized HLA-B7 Heavy Chain and Its Fragments[a]

amino acid	HLA-B7 heavy chain	h1	h2	h3
Asp[b]	25.3 (25)	8.9 (9)	10.5 (11)	5.4 (5)
Thr	16.8 (19)	4.6 (5)	3.3 (3)	10.0 (11)
Ser	14.1 (14)	7.3 (9)	1.4 (1)	2.1 (2)
Glu[b]	41.7 (42)	14.7 (15)	13.1 (13)	14.4 (14)
Pro	14.6 (15)	5.8 (6)	1.2 (1)	7.2 (8)
Gly	19.7 (19)	6.4 (6)	6.9 (7)	5.0 (5)
Ala	22.7 (23)	6.2 (6)	11.0 (11)	6.3 (6)
Cys[c]	4.3 (4)	0 (0)	1.8 (2)	1.8 (2)
Val	13.6 (12)	4.2 (4)	2.2 (2)	5.7 (6)
Met	2.3 (2)	nd[d] (1)	nd (0)	nd (0)
Ile	7.0 (7)	3.1 (3)	2.0 (2)	2.0 (2)
Leu	17.3 (17)	3.3 (3)	8.4 (8)	6.8 (6)
Tyr	12.2 (15)	7.0 (7)	5.6 (6)	1.7 (2)
Phe	6.2 (6)	3.9 (4)	0 (0)	1.9 (2)
His	9.5 (9)	1.1 (1)	1.2 (1)	5.3 (6)
Lys	7.7 (8)	1.3 (1)	3.9 (4)	3.2 (3)
Trp	8.4[e] (8)	nd (2)	nd (3)	nd (3)
Arg	25.8 (26)	10.5 (11)	9.5 (10)	4.9 (5)
residues	1-271	6-98	99-183	184-271

[a] Values are given in residues per mole. In parentheses are the number of residues as derived by sequence analysis. [b] Aspartic and glutamic acid values include asparagine and glutamine, respectively. [c] Determined as *S*-(carboxymethyl)cysteine. [d] nd = not determined. [e] Determined spectrophotometrically.

is discussed below. The amino acid composition of the HLA-B7 heavy chain and its fragments is given in Table I.

Tryptic and Chymotryptic Peptides. Fragments obtained by chemical cleavage of the polypeptide chain were used as a source material to generate both tryptic and chymotryptic peptides. Under conditions employed for tryptic digestion, most susceptible bonds were cleaved. Arginyl–proline and lysyl–proline bonds which are known to be resistant to trypsin action usually underwent partial cleavage. In the few instances in which arginyl–arginine or arginyl–lysine sequences were found [residues 145–146, 156–157, and 169–170; see the following paper (Orr et al., 1979)], heterogeneous cleavage at these positions was obtained, resulting in low recovery of

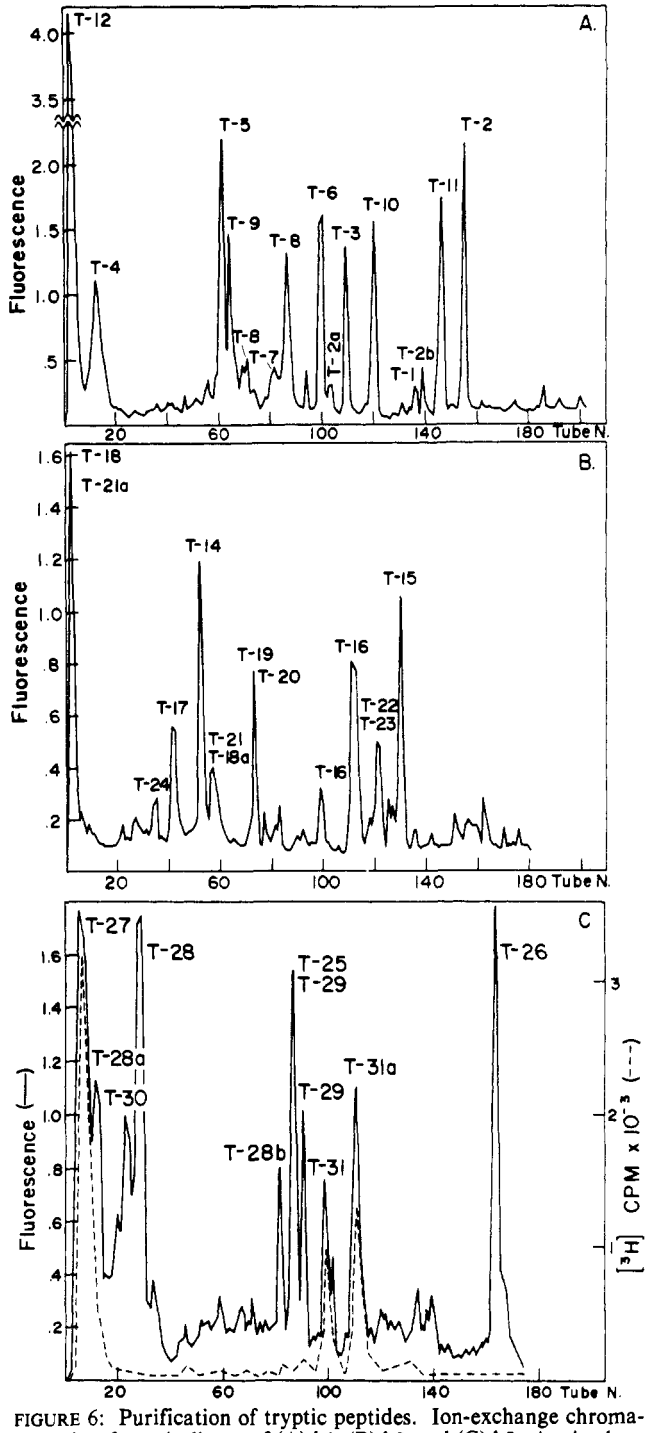

FIGURE 6: Purification of tryptic peptides. Ion-exchange chromatography of tryptic digests of (A) h1, (B) h2, and (C) h3. A microbore column (0.28 × 20 cm) was used. Elution conditions were described in the text (see Mateials and Methods). Fractions were monitored by fluorescamine reaction. The peptides identified in each peak are stated. Their numeration corresponds to the final alignment and correlates with Table II. Peptides obtained after low-yield cleavage of resistant bonds are designated by a letter (a or b) following the number of the corresponding main peptide.

Table II: Amino acid Composition of Tryptic Peptides of Papain-Solubilized HLA-B7 Heavy chain.[a]

Amino acid	T-1	T-2	T-2a	T-2b	T-3	T-4	T-5	T-6	T-7	T-8	T-9
Asp[i]					0.3(0)	2.0(2)	1.9(2)		1.1(1)	0.9(1)	1.0(1)
Thr		1.0(1)	0.9(1)			1.0(1)				0.8(1)	1.0(1)
Ser[i]	0.3(0)	1.9(2)	1.4(2)			1.0(1)	1.8(2)				
Glu[i]					0.9(1)	1.2(1)		1.9(2)	3.9(4)	1.4(1)	2.0(2)
Pro		1.1(1)		0.9(1)	1.0(1)		1.1(1)	1.0(1)	1.8(2)		
Gly		1.3(1)		1.0(1)	1.0(1)	1.4(1)			1.2(1)		
Ala									1.0(1)		1.9(2)
Cys[b]											
Val		1.1(1)	0.8(1)			2.7(3)	2.0(2)				
Met											
Ile						1.0(1)			1.0(1)	0.9(1)	
Leu											
Tyr		1.7(2)	1.8[c](2)			1.0(1)			1.0(1)	1.1(1)	
Phe		0.3(1)	1.0[c](1)			1.8(2)	1.0(1)				
His											
Lys					0.3(0)					0.9(1)	
Trp									nd[d](2)		
Arg	1.0(1)	2.1(2)	1.1(1)	1.1(1)	1.0(1)	0.9(1)	1.0(1)	1.0(1)	1.0(1)		1.0(1)
Total	1	11	8	3	4	14	9	4	14	6	7
% yield	16	19	8	19	60	24	42	49	16	37	38
Residue no.	6	7-17	7-14	15-17	18-21	22-35	36-44	45-48	49-62	63-68	69-75

Amino acid	T-10	T-11	T-12	T-13	T-14	T-15	T-16	T-17	T-18	T-19	T-20
Asp[i]		0.9(1)	1.2(1)	2.9(3)	1.8(2)		2.3(2)	2.6(3)	1.2(1)		
Thr			1.2(1)	1.0(1)					2.8(3)		
Ser[i]	0.9(1)		2.7(3)	2.8(3)					1.0(1)		
Glu[i]	1.0(1)		3.3(3)	3,5(3)			1.3(1)	1.0(1)	1.9(2)	1.3(1)	3.5(3)
Pro				1.1(1)	0.9(1)						
Gly			1.9(2)	4.9(5)	3.1(3)		1.8(2)	0.4(0)			
Ala			0.9(1)	1.3(1)			1.0(1)	1.1(1)	4.1(4)	1.8(2)	0.8(1)
Cys[b]				1.1(1)	0.4(1)						
Val				1.0(1)	1.1(1)						
Met			e(1)	0.7(1)							
Ile								0.9(1)	1.2(1)		
Leu	1.1(1)	1.2(1)	1.1(1)	1.2(1)		2.0(2)	0.4(0)	2.1(2)			
Tyr			1.4(2)	2.6(3)	1.0(1)		1.5(2)	1.2(1)			
Phe					0.3(0)						
His			1.2(1)	1.0(1)			0.8(1)				
Lys							1.2(1)				
Trp									nd[d](1)	nd[d](1)	
Arg	1.0(1)	1.0(1)		1.0(1)	1.0(1)	1.0(1)	3.2[f](0)	1.0(1)	1.1(1)	0.9(1)	1.7(2)
Total	4	3	16	26	10	3	10	10	14	5	6
% yield	63	74	59	nd	21	85	32	13	13	7	9
Residue no.	76-79	80-82	83-98	83-108	99-108	109-111	112-121	122-131	132-145	147-151	152-157

Amino acid	T-21[g]	T-22	T-23	T-24	T-25	T-26	T-27	T-28	T-29	T-30	T-31
Asp[i]	1.7(0)	1.0(1)	0.9(1)	0.9(1)		1.1(1)	0.3(0)	4.0(4)		0.4(0)	
Thr						2.7(3)	1.7(2)	2.9(3)	0.2[h](1)	0.3(0)	1.0(1)
Ser[i]	0.3(0)					1.2(1)	0.3(0)	0.4(0)		1.1(1)	
Glu[i]	3.4(3)	1.0(1)	0.8(1)	0.3(0)		1.2(1)	2.2(2)	4.9(5)	1.1(1)	3.2(3)	2.0(2)
Pro	0.4(0)				1.9(2)	1.0(1)	1.0(1)	0.9(1)		0.8(1)	1.9(2)
Gly	1.4(1)	1.2(1)				0.3(0)	1.0(1)	2.0(2)		1.4(1)	1.1(1)
Ala	1.1(1)			1.1(1)		1.0(1)	1.8(2)	1.2(1)		2.0(2)	
Cys[b]	nd (1)						0.7(1)				0.7(1)
Val	0.8(1)					0.9(1)		1.1(1)		2.5(3)	1.2(1)
Met											
Ile						1.0(1)	0.9(1)				
Leu	1.7(2)	1.0(1)	1.2(1)			1.0(1)	1.8(2)	1.2(1)		0.3(0)	2.0(2)
Tyr	nd (1)	0.8(1)					0.8(1)				0.8(1)
Phe							0.9(1)		0.9(1)		
His						3.8(4)					1.8(2)
Lys		1.0(1)	0.9(1)		1.1(1)		0.3(0)		0.4[h](1)		1.2(1)
Trp	nd[d](1)						nd[d](2)			nd[d](1)	
Arg	1.6(2)		1.2(1)			1.1(1)	0.9(1)	1.9(2)		1.0(1)	
Total	13	6	5	2	3	16	17	20	4	13	14
% yield	26	36	36	12	31	8	11	12	nd	17	6
Residue no.	158-170	171-176	177-181	182-183	184-186	187-202	203-219	220-239	240-243	244-256	257-270

(a) Values are given in residues per mol. In parenthesis are the assumed integral values, as derived by sequence analysis. (b) Determined as S-Carboxymethylcysteine. (c) Value corrected for destruction. (d) nd = not quantitatively determined. (e) Qualitatively detected as homoserine. (f) Free arginine co-elutes whith this peptide. It arises from tryptic cleavage of Arg-Arg bonds at positions 156-157 and 169-170. (g) Significant oxidation of this peptide during hydrolysis precluded the quantitation of Cys and Tyr. The presence of ASP is assumed to be a contamination since no ASP was found in sequence analysis. (h) Partially destroyed by fluorescamine reaction. (i) Aspartic and glutamic acid values include asparagine and glutamine, respectively.

the corresponding peptides (see Table II).

Fragments h1, h2, and h3 were digested with trypsin, and the resulting peptide mixtures were fractionated by microbore ion-exchange chromatography (Figure 6). Most peptides were obtained in a reasonably pure form after this single step such that they could be used for both compositional and sequence analysis. Amino acid composition of the isolated tryptic peptides is given in Table II. Peptides T-18a and T-21a (Figure 6B) were recovered in very low yield and for this reason are not included. Likewise, peptides T-28a, T-28b and

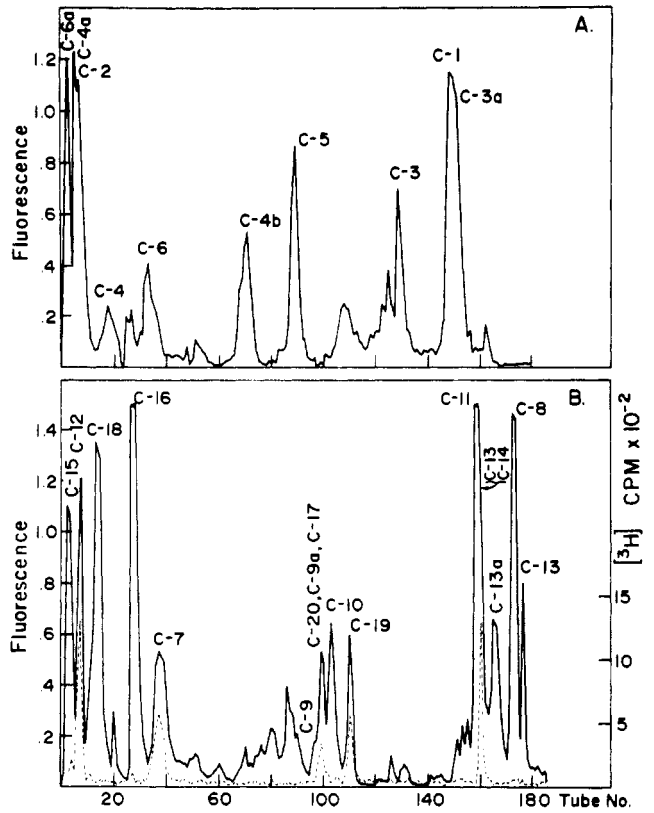

FIGURE 7: Purification of chymotryptic peptides. Microbore ion-exchange chromatography of chymotryptic digests of (A) h1 and (B) CNBr-2. Conditions were as described under Materials and Methods. Fractions were monitored by fluorescamine reaction (—). ([^{3}H]-Carboxymethyl)cysteine-containing peptides were detected by scintillation counting of small aliquots (---). Numeration of the peptides corresponds to the final alignment and correlates with Table III. Peptides obtained after low-yield cleavage of resistant bonds are designated by a letter (a or b) following the number of the corresponding main peptide.

T-31a (Figure 6C), which are products of cleavage at Arg–Pro and Lys–Pro bonds, respectively, have not been included in Table II. That cleavage at Lys_{268}–Pro_{269} bond is not due to papain is evident from the yield of chymotryptic peptides C-19 and C-20 which extend beyond Lys_{268}.

The conditions employed for chymotryptic digestion largely restricted the enzymatic action to peptide bonds at which aromatic residues were involved. The enzyme did not cleave at leucine bonds or did so with low yield, the only significant exception being leucyl–arginine peptide bonds which were frequently hydrolyzed. This resulted in a limited number of peptides which were purified in preparative amounts by ion-exchange chromatography (Figure 7). Table III presents the amino acid composition of the major chymotryptic peptides generated. Some other peptides were found which resulted from low-yield cleavage at several points (peptides C-3a, C-6a, C-9a, and C-13a; see Figure 7). These were recovered in very small amounts and for this reason were not included in Table III.

Regardless of the amount of starting material, neither tryptic nor chymotryptic digestion of any of the fragments generated detectable amounts of insoluble peptides. Thus, the peptides obtained by ion-exchange chromatography and characterized in Tables II and III account for the complete extent of the papain-solubilized HLA-B7 heavy chain. The only exception is the amino-terminal pentapeptide generated by CNBr cleavage at methionine-5. The sequence of the amino-terminal region was determined directly by automated Edman degradation of the purified heavy chain (Orr et al., 1979). Thus,

Table III: Amino acid composition of chymotryptic peptides of papain-solubilized HLA-B7 heavy chain.[a]

Amino acid	C-1	C-2	C-3	C-4	C-4a	C-4b	C-5	C-6	C-7	C-8	C-9	C-10	C-11	C-12	C-13	C-14	C-15	C-16	C-17[f]	C-18	C-19	C-20
Asp[g]	0.4(0)	2.0(2)	2.5(2)	1.8(2)	0.3(0)	2.1(2)	1.2(1)	2.1(2)	1.9(2)	1.0(1)	3.3(4)	1.0(1)			3.7(4)		0.3(0)	3.5(4)			0.5(0)	
Thr	1.2(1)	1.1(1)	0.4(0)	0.9(1)		0.9(1)	0.9(1)	1.0(1)			0.5(0)	2.7(3)			2.8(3)		1.8(2)	3.4(4)			1.0(1)	1.9(2)
Ser	3.2(3)	0.4(0)	2.3(2)	0.5(0)	0.5(0)	0.3(0)	0.9(1)	2.5(3)			1.0(1)				1.0(1)			0.4(0)		1.0(1)		
Glu[g]	1.2(1)	1.2(1)	2.4(2)	5.5(5)	4.0(4)	1.2(1)	3.0(3)	3.1(3)		1.0(1)	1.4(1)	2.0(2)	4.0(4)	2.8(3)	3.3(3)		1.3(1)	6.0(6)	(1)	3.1(3)	2.0(2)	1.9(2)
Pro	1.8(2)		2.7(3)	0.9(1)	0.7(1)				1.1(1)						3.0(3)		1.0(1)	1.0(1)		1.1(1)	1.8(2)	2.1(2)
Gly	3.1(3)	0.3(0)	0.8(0)	1.3(1)	1.1(1)			2.8(2)	3.0(3)	0.9(1)	1.2(1)			1.0(1)	1.1(1)		1.1(1)	3.8[e](2)		1.1(1)	1.2(1)	1.1(1)
Ala[c]	0.4(0)		2.6(3)	0.3(0)		0.5(0)	1.9(2)	1.3(1)		0.3(0)	1.2(1)	3.8(4)	4.0(4)		2.2(2)		2.2(2)	0.9(1)		2.1(2)	0.4(0)	0.6(0)
Cys									0.7(1)					0.7(1)		1.0(1)					0.7(1)	0.8(1)
Val[d]	2.1(2)	0.9(1)	1.2(1)						0.9(1)					1.0(1)	0.9(1)			0.9(1)		2.7(3)	0.9(1)	0.9(1)
Met								(1)														
Ile	1.1(1)			1.7(2)	0.9(1)	0.9(1)					1.1(1)	1.2(1)			1.2(1)		1.0(1)	1.2(1)				
Leu		0.3(0)	0.6(0)		0.3(0)		1.1(1)	2.0(2)	2.3(2)		2.0(2)	0.3(0)		1.2(1)	3.1(3)		2.0(2)				2.3(2)	1.8(2)
Tyr					0.3(0)			1.5(2)	0.8(1)	1.2(1)	0.8(1)		1.0(1)				0.8(1)			1.0(1)		
Phe	0.8(1)	0.8(1)	0.8(1)	1.9(2)	1.2(1)	0.8(1)											0.9(1)	0.7(1)				
His	0.8(1)							0.8(1)		1.0(1)			0.3(0)		3.7(4)						1.8(2)	2.0(2)
Lys	0.4(0)		0.3(0)				1.0(1)		1.0(1)		1.0(1)	1.1(1)			2.8(3)			0.5(0)	(1)		1.0(1)	1.0(1)
Trp			nd[b](1)	nd (1)	nd (1)				nd (1)		nd (1)	nd (1)		nd (1)		nd (1)	nd (1)		(1)			
Arg	2.7(3)		2.6(3)	0.9(1)		1.1(1)	1.0(1)	1.9(2)	1.1(1)	0.8(1)	1.0(1)	1.1(1)	3.0(3)		1.3(1)	1.0(1)		2.6(3)		1.0(1)	0.3(0)	0.3(0)
Total	18	6	18	16	9	7	11	20	12	6	15	14	12	8	30	3	13	24	3	13	13	14
I yield	11	27	13	18	22	11	29	10	13	46	9	14	23	12	14	73	40	26	nd	42	19	18
Residue no.	1G-27	28-33	34-51	52-67	52-60	61-67	68-78	79-98	99-110	111-116	119-133	134-147	148-159	160-167	172-201	202-204	205-217	218-241	242-244	245-257	258-270	258-271

(a) Values are given in residues per mol. in parentheses are the values as derived by sequence analysis.
(b) nd= Not quantitatively determined.
(c) Detected as S-carboxymethylcysteine.
(d) Detected as homoserine.
(e) This value reflects the presence of some contamination. No evidence for additional Gly was found in the composition of the tryptic peptides T-27, T-28 and T-29 covering this region (see Table II) or by sequence analysis.
(f) Composition of this peptide is given as derived by sequence analysis.
(g) Aspartic and glutamic acid values include asparagine and glutamine, respectively.

recovery and further characterization of the fragment were not pursued.

Discussion

Structural characterization of human histocompatibility antigens has been underway for a number of years. Fundamental information concerning their quaternary structure (Cresswell et al., 1973, 1974a), their nature as integral membrane glycoproteins (Springer & Strominger, 1976), and the linear arrangement (Peterson et al., 1975; Terhorst et al., 1977) and size of the intrachain disulfide bonds (Ferguson et al., 1979) has been obtained. However, a complete analysis of the molecule at the level of primary structure has been made possible only after the purification of sizable amounts of homogeneous material. This was the result of the following: (1) the establishment of a transformed human lymphoblastoid cell line, JY, homozygous at both HLA-A and -B loci (HLA-A2,2 and -B7,7); this line is stable and suitable for large-scale in vitro culture; (2) solubilization of a major portion of the histocompatibility antigen molecule by treatment of membranes with papain (Mann et al., 1968); (3) preparative-scale purification of single specificities of papain-solubilized material by chromatographic procedures (Turner et al., 1975; Parham et al., 1977).

The strategy for the primary structure determination of the papain-solubilized portion of HLA-B7 heavy chain centered on the presence of an aspartyl–proline bond between the two disulfide bridges of the molecule (Terhorst et al., 1977), which provided a selective cleavage point by partial acid hydrolysis (Piszkiewizk et al., 1971), and of two methionine residues, one near the amino terminus (Terhorst et al., 1976) and the other located between the carbohydrate moiety and the first half-cystine residue of the molecule (Terhorst et al., 1977) which provided additional points for specific CNBr cleavage. Thus, three large fragments, h1, h2, and h3, could be obtained. The strategic location of the aspartyl–proline bond made the separation of the acid cleavage fragments by gel filtration a simple matter. The larger fragment (ac-1) could be further cleaved by CNBr, generating two fragments, h1 and h2. The yield of this reaction was very dependent upon the presence of a reducing agent during the preparation of ac-1. It is well-known that oxidation of the thioether group at the methionine side chain makes this residue resistant to CNBr attack (Kasper, 1970). Oxidative conditions are likely to occur during the acid cleavage reaction in formic acid due to the presence of oxygen or trace amounts of peroxides, unless a reducing agent is added. Fragments h1 and h2 were easily separated by gel filtration due to the presence of the carbohydrate moiety in h1, which significantly added to the size of the fragment. Thus, all three fragments, h1, h2, and h3, could be purified by gel filtration.

An alternative strategy was to initially cleave with CNBr. The two major fragments produced were readily soluble in bicarbonate buffers and could be separated by ion-exchange chromatography. This procedure has several advantages: (1) fragments suitable for protein sequence analysis may be obtained in higher yield; (2) the harsh conditions involved in the acid cleavage reaction are avoided; these conditions may produce significant deamidation, making it difficult to identify the amides (Asn, Gln); (3) the different cleavage pattern provides the overlap for the acid cleavage derived fragments which is necessary to formally establish that no residue or small peptide is lost during the acid cleavage.

The alignment of the fragments and a schematic model for the papain-solubilized HLA-B7 heavy chain is given in Figure 5. The alignment has been discussed on the basis of preliminary data (Terhorst et al., 1977; Trägårdh et al., 1979). Our results confirm these previous reports. The continuity of the polypeptide chain is formally established at the methionine residue near the NH_2 terminus through the N-terminal amino acid sequence (Terhorst et al., 1976; Orr et al., 1979). The amino acid composition of tryptic glycopeptide T-13 (Table II) obtained after tryptic digestion of the complete HLA-B7 heavy chain (Parham et al., 1977) contains a single methionine residue and its amino acid composition is equal to the sum of peptides T-12 and T-14, obtained from fragments h1 and h2, respectively, thus establishing the continuity between these two fragments. Overlap between the acid cleavage derived fragments is provided by the chymotryptic peptide C-13 (Table III) obtained from the CNBr fragment CNBr-2. This peptide contains the only aspartyl–proline bond of the molecule as judged by sequence analysis (Orr et al., 1979).

As shown in Figure 5, the chemical cleavage of HLA-B7 heavy chain dissects the molecule into three fragments of similar size. Fragment h1 contains the single carbohydrate moiety (Parham et al., 1977) and does not contain any cysteine residues. This fragment extends from residues 6 to 98, as determined by sequence analysis (Orr et al., 1979). Fragment h2 contains the first disulfide bond of the molecule. Peptide and sequence analyses show that it extends from residues 99 to 183. Fragment h3 contains the second disulfide bond and extends from residue 184 to the heterogeneous COOH terminus of the molecule at positions 270 and 271.

Given earlier speculation that the heavy chain of histocompatibility antigens may consist of independent domains in analogy with the organization of the antibody molecule (Strominger et al., 1974; Peterson et al., 1975; Terhorst et al., 1977), it is tempting to correlate these fragments with independent domains. Therefore, we should stress that the existence of peptide bonds susceptible to specific chemical cleavage is a fortuitous event with no structural or functional significance, whatever its practical importance from a methodological point of view. A complete discussion of this point is presented in the following paper (Orr et al., 1979).

A final word should be said regarding methodology. Given the limitations imposed by the high cost and limited supply of material, purification and sequencing of peptides were worked out in the nanomole range. Ion-exchange chromatography on microbore columns provided an adequate method of peptide purification for preparative purposes at this level because of the relative high recovery and high resolution of the chromatography. In spite of these advantages, very complex peptide mixtures resulted in significant coelutions, limiting the size of the fragments which could be used for enzymatic digestions. Attempts of further purification resulted in a significant decrease of yield.

Because of the limited size of h1, h2, and h3 fragments, peptide mixtures were obtained after enzymatic digestion which allowed preparative-scale purification of most peptides in a single step. In this way, tryptic and chymotryptic peptides covering the whole extent of the papain-solubilized HLA-B7 heavy chain were obtained and characterized, with the exception of the five amino-terminal residues as mentioned earlier. The complete amino acid sequence of the molecule was obtained from the combined sequence analysis of these peptides and the amino-terminal automatic sequencing of the whole heavy chain and its fragments, as described in the following paper.

References

Cresswell, P., Turner, M. J., & Strominger, J. L. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 1603–1607.

Cresswell, P., Springer, T., Strominger, J. L., Turner, M. J., Grey, H. M., & Kubo, R. T. (1974a) *Proc. Natl. Acad. Sci. U.S.A. 71*, 2123–2127.
Cresswell, P., Robb, R. J., Turner, M. J., & Strominger, J. L. (1974b) *J. Biol. Chem. 249*, 2828–2832.
Dickmeiss, E., Soeberg, B., & Svejgaard, A. (1977) *Nature (London) 270*, 526–528.
Ferguson, W. S., Terhorst, C. T., Robb, R. J., & Strominger, J. L. (1979) *Mol. Immunol. 16*, 23–28.
Gracy, R. W. (1977) *Methods Enzymol. 47*, 195–204.
Gross, E. (1967) *Methods Enzymol. 11*, 238–255.
Herman, V. A., & Vanaman, T. C. (1977) *Methods Enzymol. 47*, 220–236.
Kasper, C. B. (1970) in *Protein Sequence Determination* (Needleman, S. B., Ed.) pp 137–184, Springer-Verlag, West Berlin.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Lai, C. Y. (1977) *Methods Enzymol. 47*, 236–243.
Machleidt, W., Otto, J., & Wachter, E. (1977) *Methods Enzymol. 47*, 210–220.
Maizel, J. V., Jr. (1971) *Methods Virol. 5*, 180–247.
Mann, D. L., Rogentine, Y. N., Jr., Fahey, J. L., & Nathenson, S. G. (1968) *Nature (London) 217*, 1180–1181.
McMichael, A. J., Ting, A., Zweerink, H. J., & Askonas, B. (1977) *Nature (London) 270*, 524–526.
Orr, H. T., Lôpez de Castro, J. A., Lancet, D., & Strominger, J. L. (1979) *Biochemistry* (following paper in this issue).
Parham, P., Alpert, B. N., Orr, H. T., & Strominger, J. L. (1977) *J. Biol. Chem. 252*, 7555–7567.
Paul, W. E., & Benacerraf, B. (1977) *Science 195*, 1293–1300.
Peterson, P. A., Rask, L., Sege, K., Klareskog, L., Anundi, H., & Ostberg, L. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1612–1616.
Piszkiewizk, D., Landon, M., & Smith, E. L. (1971) *Biochem. Biophys. Res. Commun. 40*, 1173–1178.
Poulik, M. D., & Reisfeld, R. A. (1976) *Contemp. Top. Immunobiol. 4*, 157–204.
Sanderson, A. (1968) *Nature (London) 220*, 192–195.
Springer, T. A., & Strominger, J. L. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 2481–2485.
Strominger, J. L., Cresswell, P., Grey, H., Humphreys, R. E., Mann, D., McCune, J., Parham, P., Robb, R., Sanderson, A. R., Springer, T. A., Terhorst, C., & Turner, M. J. (1974) *Transplant. Rev. 21*, 126–143.
Terhorst, C., Parham, P., Mann, D. L., & Strominger, J. L. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 910–914.
Terhorst, C., Robb, R., Jones, C., & Strominger, J. L. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4002–4006.
Thorsby, E. (1978) *Tissue Antigens 11*, 321–329.
Trägårdh, L., Wiman, K., Rask, L., & Peterson, P. A. (1979) *Biochemistry 18*, 1322–1328.
Turner, M. J., Cresswell, P., Parham, P., & Strominger, J. L. (1975) *J. Biol. Chem. 250*, 4512–4519.
Tursz, T., Fridman, W. H., Senik, A., Tsapis, A., & Fellous, M. (1977) *Nature (London) 269*, 806–808.
van Someran, H., Westerveld, A., Hagemeijer, A., Mees, J. R., Kahan, M., & Zaalberg, O. B. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 962–965.

Complete Amino Acid Sequence of a Papain-Solubilized Human Histocompatibility Antigen, HLA-B7. 2. Sequence Determination and Search for Homologies†

Harry T. Orr, Jośe A. Lôpez de Castro, Doron Lancet, and Jack L. Strominger*

ABSTRACT: The complete amino acid sequence of the papain-solubilized heavy chain of a human histocompatibility antigen (HLA-B7) has been elucidated. It consists of a polypeptide of 271 residues (31 333 daltons). A single glycan moiety is attached to an asparagine residue at position 86 by an N-glycosidic bond. Two intrachain disulfide bonds, arranged linearly, involve half-cystine residues at positions 101 and 164 and at positions 203 and 259. They form two loops of 62 and 55 residues, respectively, separated by 38 residues. Computer analysis of the sequence suggests the existence of internal homology between the amino-terminal portion (residues 1–90) and the region of the first disulfide loop (residues 91–180). There is a significant homology between the second disulfide loop region of the chain (residues 182–271) and immunoglobulin (Ig) constant domains and β_2-microglobulin [Orr, H. T., Lancet, D., Robb, R. J., Lôpez de Castro, J. A., & Strominger, J. L. (1979a) *Nature (London)* (in press)]. However, no such homology to Ig is apparent in the amino-terminal or in the first disulfide loop regions.

A relationship between the highly polymorphic HLA-A and HLA-B antigens and immunoglobulins has been suggested as the consequence of several properties of these antigens. β_2-Microglobulin, an immunoglobulin-like polypeptide (Peterson et al., 1972; Smithies & Poulik, 1972), is one of the two components of the HLA-A and -B molecule (Cresswell et al., 1973, 1974; Nakamuro et al., 1973; Grey et al., 1973; Peterson et al., 1974). The presence in the heavy chain of these antigens of two disulfide loops similar in arrangement and size to those in immunoglobulins added credibility to this speculation (Strominger et al., 1974; Peterson et al., 1975; Terhorst et al., 1977; Ferguson et al., 1979). Subsequently, the amino acid sequence of the second disulfide loop region of the HLA-B7 heavy chain has provided strong evidence that this region of a histocompatibility antigen is related to an Ig[1] constant do-

† From the Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138. *Received July 17, 1979.* Supported by a research grant from the National Institutes of Health (AI 10736). H.T.O. has a U.S. Public Health Service Research Fellowship (AI 05483), and J.A.L.C. has a U.S. Public Health Service International Research Fellowship (TW 02582). D.L. has a Dr. Chaim Weizmann Postdoctoral Fellowship for Scientific Research.

0006-2960/79/0418-5711$01.00/0