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

main (Orr et al., 1979a). Obviously, a detailed knowledge of the chemical structure of the histocompatibility antigen molecule is essential both to define the molecular basis of the genetic polymorphism and to determine the extent of the relationship with the immunoglobulin system.

The existence of genetically defined strains of mice and the development of methods for the metabolic incorporation of virtually all radioactive amino acids have made possible extensive sequence studies of murine histocompatibility antigens (Coligan et al., 1979). A more conventional approach has been used in the elucidation of the complete amino acid sequence of papain-solubilized human histocompatibility antigen (HLA-B7) heavy chain presented in this paper. The sequence data have been analyzed by statistical methods to assess similarities to other proteins, as well as possible internal homologies. The results are discussed in terms of the overall structure of HLA-B7 and its relation to immunoglobulins.

Materials and Methods

The isolation of papain-solubilized HLA-B7 heavy chain and its chemical and proteolytic cleavage are described in the preceding paper (López de Castro et al., 1979), except for the cleavage at tryptophan residues. Such cleavage was performed according to Mahoney & Hermodson (1979). Briefly, 5 mg of HLA-B7$_{pap}$ heavy chain was dissolved in a 4 M solution of guanidine hydrochloride in 80% (v/v) acetic acid–H_2O, and then 10 mg of *o*-iodosobenzoic acid, pretreated with 0.1 mol of *p*-cresol per mol of iodosobenzoic acid for 2 h in the dark at room temperature, was added to the protein and the mixture was allowed to react for 24 h in the dark at room temperature. The fragments generated were separated on a Sephadex G-50 column (0.9 × 200 cm) in 200 mM NH_4HCO_3.

Carboxypeptidase Analysis. The carboxy-terminal residues were determined by amino acid analysis after carboxypeptidase A digestion of papain-solubilized HLA-B7 heavy chain (Ambler, 1967). Digestion was performed in 20 mM *N*-ethylmorpholine at 25 °C. At various times, aliquots were removed and acid precipitated and the supernatants were lyophilized and analyzed in a Beckman 121M automatic amino acid analyzer. Aliquots were taken, and protein was determined after acid hydrolysis. All results were normalized to an added internal standard of norleucine.

Sequence Determination. All large fragments and most peptides were sequenced by using an updated Beckman 890 sequenator. A few peptides, generally those of less than five residues, were sequenced by using a modification of the three-step manual Edman method (Sauer et al., 1974). In all automatic degradations, 3 mg of Polybrene (Tarr et al., 1978; Klapper et al., 1978) was added to the cup and a complete blank cycle run before the addition of the peptide sample. The 0.1 M Quadrol program (Brauer et al., 1975) with double coupling at the first step was used for most degradations. To prevent Quadrol retention in the sequencing of small glycopeptides, DMAA buffer (11 mL of pyridine, 11 mL of H_2O, 2.8 mL of DMAA, and 50 μL of trifluoroacetic acid) was substituted as the coupling buffer. The remainder of the program was identical with the 0.1 M Quadrol program. Repetitive yields varied from 90.7 to 94% (see Table II). In all cases, the yield at step 1 was between 70 and 80% of the total amount added to the cup.

Pth Determination. Conversion of the anilinethiazolinone fractions to the corresponding phenylthiohydantoin (Pth)–amino acid derivatives was performed with 0.1 N HCl at 80 °C for 10 min. Pth derivatives were identified by gas–liquid chromatography (Pisano & Bronzert, 1969), two-dimensional thin-layer chromatography (Summers et al., 1973; Kulbe, 1974), and back hydrolysis followed by amino acid analysis (Smithies et al., 1971). Cysteine was identified as Pth-*S*-(carboxymethyl)cysteine. Pth-leucine and Pth-isoleucine were distinguished either by silylation using *N,O*-bis[(trimethylsilyl)acetamide] and gas–liquid chromatography on a SP-400 column (Niall, 1973) or by back hydrolysis (Smithies et al., 1971). Pth-arginine and Pth-histidine were identified by amino acid analysis after back hydrolysis.

Reagents. The following commercial sources were used to obtain reagents: [^{3}H]iodoacetic acid (New England Nuclear); polyamide thin-layer plates, Polybrene, *o*-iodosobenzoic acid, and DMAA (Pierce Chemical); benzene, 1-chlorobutane, heptane, heptafluorobutyric acid, phenyl isothiocyanate, 1,2-cyclohexadiene, ethyl acetate, and Quadrol buffer (Beckman Instruments); pyridine and hydriodic acid (Fisher Scientific). DMAA was refluxed with phthalic anhydride and pyridine with ninhydrin. Both were then redistilled twice before use in the DMAA buffer.

Computer Analysis. Two computer programs have been used. The first, IH04, is analogous to SEARCH (Dayhoff, 1976) and compares a segment of a sequence to all possible segments of the same length, in the same sequence or in other sequences, looking for best alignments without gaps. The elements of the mutation data matrix [cf. Dayhoff et al. (1972); revised values taken from Schwartz & Dayhoff (1978)] were used to assess pairwise matches of amino acids, and those were summed to yield the alignment score. The second program, AL04, is equivalent to ALIGN (Dayhoff, 1976) and is described in Orr et al. (1979a). It is based on the algorithm of Needleman & Wunsch (1970), locating the best alignments between two sequences of comparable size, with gaps introduced in both. The same mutation data matrix is used to assess pairwise residue matches. Gap introduction is controlled and standardized by the method of Dayhoff (1976). A gap penalty of −2 and a bias parameter of 2 were used unless otherwise mentioned. The statistical scores (*Z* factors) were calculated as described in Barker & Dayhoff (1972). In short, alignment scores (see above) were calculated for the two sequences in question and then, using the same gap control parameters, also for one of the sequences with many random permutations of the other. The statistical score is defined as $Z = (s - \mu)/\sigma$, where s is the score for the actual proteins and μ and σ are the mean and standard deviation, respectively, for the distribution of scores for randomized sequences. Thus, Z is a measure, in standard deviation units (SD), of how likely any particular value of s is to arise by chance (e.g., Z equal to 3 and 6 corresponds to such probability of 10^{-3} and 10^{-9}, respectively).

Results

NH_2-Terminal Sequence Analysis of HLA-B7 Heavy Chain. Automatic Edman degradation of HLA-B7 heavy chain was carried out for 26 cycles (Table I). A single methionine residue was identified at position 5. Thus, this sequence extends through the amino-terminal CNBr pentapeptide and into the h1 region. Previous sequence data from this portion of the HLA-B7 heavy chain are very similar to those reported here (Terhorst et al., 1976), the exceptions being residues 21 and 24, where Terhorst et al. (1976) found Pro and Ala, respectively. These discrepancies are likely due to

[1] Abbreviations used: Ig, immunoglobulin; Ig C, constant domains; Ig V, variable domains; DMAA, *N,N*-dimethyl-*N*-allylamine; Pth, phenylthiohydantoin; CNBr, cyanogen bromide; HLA-B7$_{pap}$, HLA-B7 heavy chain solubilized after papain digestion; β_2m, β_2-microglobulin.

Table I: N-Terminal Sequence of HLA-B7 Heavy Chain and Its Fragments[d]

HLA-B7 heavy chain	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 Gly-Ser-His-Ser-Met-Arg-Tyr-Phe-Tyr-Thr-Ser-Val-Ser-Arg-Pro-Gly-Arg-Gly-Glu-Pro-Arg-Phe-Ile-Ser-Val-Gly... 15 *a* 10 *a* 16 8 7 13 7 *b* *a* 14 *a* 5 6 11 4 8 6 4 3 5 9 *a* 6 3
CNBr-2	99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 Tyr-Gly-Cys-Asp-Val-Gly-Pro-Asp-Gly-Arg-Leu-Leu-Arg-Gly-His-Asp-Gln-Tyr-Ala-Tyr-Asp-Gly-Lys-Asp-Tyr-Ile- 31 22 18 20 29 18 10 12 11 8 14 13 7 10 7 9 11 12 9 10 8 6 7 6 7 8
	125 126 127 128 129 130 131 132 133 134 135 136 Ala-Leu-Asn-Gln- -Leu- - -Trp-Thr-Ala-Ala... 7 5 4 3 3 *c* *b* 2 1
h3 (ac-2)	184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 Pro-Pro-Lys-Thr-His-Val-Thr-His-His-Pro-Ile-Ser-Asp-His-Glu-Ala-Thr-Leu-Arg-Cys-Trp-Ala-Leu-Gly-Phe... 8 5 6 *b* 4 6 *b* 3 2 3 3 *a* 2 1 2 2 *b* 2 1 1 *c* 2 1 1 1
I-6	168 169 170 171 172 173 174 175 176 177 Leu-Arg-Arg-Tyr-Leu-Glu-Asn-Gly-Lys-Asp... 7 2 3 6 5 *c* *c* 2 *c* *c*

[a] Identified as the Pth derivative by gas-liquid chromatography but not quantitated. [b] Pth derivative identified by gas-liquid chromatography, thin-layer chromatography, and after back hydrolysis but not quantitated. [c] Denotes Pth derivatives identified only by thin-layer chromatography. [d] Numbers under residues are yields in nanomoles of the Pth derivatives. Yields were calculated from GC data or, in the case of Pth-Arg and Pth-His, from amino acid analysis after back hydrolysis. Numbers above residues indicate the position in the HLA-$B7_{pap}$ molecule. Peptides sequenced to the carboxyl termini end with COOH. Peptides and fragments which were not sequenced completely end with (...). Repetitive yields were 93.3% for the HLA-B7 heavy chain, 94% for CNBr-2, and 90.7% for h3 (ac-2).

Table II: Sequence of Peptides from the h1 Fragment of HLA-B7 Heavy Chain[d]

T-4	22 23 24 25 26 27 28 29 30 31 32 33 34 35 Phe-Ile-Ser-Val-Gly-Tyr-Val-Asp-Asp-Thr-Gln-Phe-Val-Arg-COOH 28 42 *a* 45 19 24 22 18 12 *b* 10 7 4 3
T-5	36 37 38 39 40 41 42 43 44 Phe-Asp-Ser-Asp-Ala-Ala-Ser-Pro-Arg-COOH 19 20 *a* 18 16 15 *a* 6 5
T-6	45 46 47 48 Glu-Glu-Pro-Arg-COOH 25 22 9 8
T-7	49 50 51 52 53 54 55 56 57 58 59 60 61 62 Ala-Pro-Trp-Ile-Glu-Gln-Glu-Gly-Pro-Glu-Tyr-Trp-Asp-Arg-COOH 8 3 *c* 8 3 3 3 4 1 2 1 *c* 1 1
T-8	63 64 65 66 67 68 Asn-Thr-Gln-Ile-Tyr-Lys-COOH 5 *b* 4 6 3 2
T-9	69 70 71 72 73 74 75 Ala-Gln-Ala-Gln-Thr-Asp-Arg-COOH 45 32 24 21 *b* 20 15
T-10	76 77 78 79 Glu-Ser-Leu-Arg-COOH 22 *a* 17 20
T-11	80 81 82 Asn-Leu-Arg-COOH 45 41 32
T-12	83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 Gly-Tyr-Tyr-Asn-Gln-Ser-Glu-Ala-Gly-Ser-His-Thr-Leu-Gln-Ser... 30 20 16 7[e] 14 *a* 12 11 7 *a* 5 *b* 4 *c* *a*
C-3	34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 Val-Arg-Phe-Asp-Ser-Asp-Ala-Ala-Ser-Pro-Arg-Glu-Glu-Pro-Arg-Ala-Pro-Trp-COOH 6 4 6 6 *a* 5 3 3 *a* 1 1 0.5 0.5 0.5 0.5 1 *a* *c*
C-4b	61 62 63 64 65 66 67 Asp-Arg-Asn-Thr-Gln-Ile-Tyr-COOH 1 0.5 *c* *b* *c* 1 1
C-5	68 69 70 71 72 73 74 75 76 77 78 Lys-Ala-Gln-Ala-Gln-Thr-Asp-Arg-Glu-Ser-Leu-COOH 10 9 6 6 5 *b* 1 1 0.5 *a* 1
C-6	79 80 81 82 83 84 85 86 87 Arg-Asn-Leu-Arg-Gly-Tyr-Tyr-Asn-Gln... 1 1 2 1 2 2 1 1 *c*

[a-d] See footnotes *a-d* in Table I. [e] Yield calculated from amino acid analysis after back hydrolysis of both organic and aqueous phases.

the use of a nonquantitative identification method in the earlier study.

Sequence of h1. This fragment extends from Arg_6 to Met_{98}. The sequence of tryptic peptides isolated from h1 [see Lôpez de Castro et al. (1979)] is depicted in Table II. Peptides T-1, T-2 (T-2a, T-2b), T-3, and T-4 were characterized by amino acid composition [see Lôpez de Castro et al. (1979)] and aligned by the amino-terminal sequence of HLA-B7 heavy chain. Peptide T-12 was the only tryptic peptide which did not contain Arg or Lys. Its amino acid composition included homoserine, and therefore it was assigned as the carboxy-terminal peptide of the fragment. Edman degradation of chymotryptic peptides C-3, C-4b, C-5, and C-6 (Table II) provided the remaining overlaps required to establish the complete sequence of h1.

Fragment h1 is the only large fragment of papain-solubilized HLA-B7 heavy chain which does not contain a disulfide loop. It contains the single glycan moiety of HLA-B7, attached to Asn_{86} through an N-glycosidic bond (Parham et al., 1977). Unlike mouse $H\text{-}2K^b$, there is no evidence for a second car-

Table III: Sequence of Tryptic Peptides from h2 and h3 Fragments of HLA-B7 Heavy Chain[d]

T-16	112 113 114 115 116 117 118 119 120 121 Gly- -Asp-Gln-Tyr-Ala-Tyr-Asp-Gly-Lys-COOH *c* *c* *c* *c* *c* *c* *c* *c*
T-20	152 153 154 155 Glu-Ala-Glu-Gln... 9 8 5 4
T-21	158 159 160 161 162 163 164 165 166 167 168 169 170 Ala-Tyr-Leu-Glu-Gly-Glu-Cys-Val-Glu-Trp-Leu-Arg-Arg-COOH 8 6 6 4 3 5 2 4 2 *c* 1 1
T-22	171 172 173 Tyr-Leu-Glu... 3 2 2
T-23	177 178 179 180 Asp-Lys-Leu-Glu... 6 4 5 2
T-26	187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 Thr-His-Val-Thr-His-His-Pro-Ile-Ser-Asp-His-Glu-Ala-Thr-Leu-Arg-COOH *b* 3 6 *b* 3 2 1 4 *a* 1 0.5 1 3 *b* 1 1
T-27	203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 Cys-Trp-Ala-Leu-Gly-Phe-Tyr-Pro-Ala-Glu-Ile-Thr-Leu-Thr-Trp-Gln-Arg-COOH *b* *c* 5 6 5 2 3 2 4 2 4 *b* 3 *b* *c* 1 0.5
T-28	220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 Asp-Gly-Glu-Asp-Gln-Thr-Gln-Asp-Thr-Glu-Leu-Val-Glu-Thr-Arg-Pro-Ala-Gly-Asp-Arg-COOH 3 4 2 2 2 *b* 1 1 *b* 1 1 1 1 *c* 0.5 0.2 1 0.5 *c* 0.2
T-30	244 245 246 247 248 249 250 251 252 253 254 255 256 Trp-Ala-Ala-Val-Val-Val-Pro-Ser-Gly-Glu-Glu-Gln-Arg-COOH *c* 9 8 9 9 6 2 *a* 4 3 2 1 1
T-31	257 258 259 260 261 262 263 264 265 266 267 Tyr-Thr-Cys-His-Val-Gln-His-Glu-Gly-Leu-Pro... 4 *b* *b* 2 4 3 1 2 2 3 *a*

[a-d] See footnotes *a-d* in Table I.

Table IV: Sequence of Chymotryptic Peptides from the h2,3 Fragment of HLA-B7 Heavy Chain[d]

C-9	119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 Asp-Gly-Lys-Asp-Tyr-Ile-Ala-Leu-Asn-Glu-Asp-Leu-Arg-Ser-Trp-COOH 3 2 3 2 3 4 3 4 2 2 1 2 1 *a* 1
C-10	134 135 136 137 138 139 140 141 142 143 144 145 146 147 Thr-Ala-Ala-Asp-Thr-Ala-Ala-Gln-Ile-Thr-Gln-Arg-Lys-Trp-COOH *b* 9 7 6 *b* 8 7 5 7 *b* 3 2 1 1
C-11	148 149 150 151 152 153 154 155 156 157 158 159 Glu-Ala-Ala-Arg-Glu-Ala-Glu-Gln-Arg-Arg-Ala-Tyr-COOH 20 16 17 6 8 10 6 5 5 4 3 3
C-13	172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 Leu-Glu-Asn-Gly-Lys-Asp-Lys-Leu-Glu-Arg-Ala-Asp-Pro-Pro-Lys-Thr... 4 3 4 2 3 4 3 2 1 1 2 1 1 0.5 *c* *c*
C-14	202 203 204 Arg-Cys-Trp-COOH 8 5 *c*
C-15	205 206 207 208 209 210 211 212 213 214 215 216 217 Ala-Leu-Gly-Phe-Tyr-Pro-Ala-Glu-Ile-Thr-Leu-Thr-Trp-COOH 14 13 13 11 10 4 9 5 7 *b* 3 *b* *c*
C-17	242 243 244 Glu-Lys-Trp-COOH 8 6 *c*
C-18	245 246 247 248 249 250 251 252 253 254 255 256 257 Ala-Ala-Val-Val-Val-Pro-Ser-Gly-Glu-Glu-Gln-Arg-Tyr-COOH 18 15 11 15 14 4 *a* 11 6 6 5 2 2
C-19	258 259 260 261 262 263 264 265 266 267 268 Thr-Cys-His-Val-Gln-His-Glu-Gly-Leu-Pro-Lys... *b* *b* 5 9 4 3 2 3 2 1 *c*

[a-d] See footnotes *a-d* in Table I.

bohydrate moiety in HLA-B7 heavy chain. In fact, the tripeptide Asn-Gln-Ser is the only one in the molecule which fits the general sequence of Asn-X-Ser/Thr common among N-linked glycosylation sites (Kronquist & Lennarz, 1978).

Sequence of h2 and h3. This portion of the molecule extends from Tyr_{99} to Thr_{271}. Fragments CNBr-2 and h3 provided starting points for automatic Edman degradation at positions 99 and 184, respectively. The results of these degradations are given in Table I. In addition, tryptic peptides were obtained from the h2 and h3 fragments and chymotryptic peptides were obtained from the CNBr-2 fragment [see Lôpez de Castro et al. (1979)]. The sequence of some of these peptides is given in Tables III and IV. Peptides T-14–T-18 and C-7–C-10 were aligned based on the amino-terminal sequence of CNBr-2 and their amino acid compositions (Lôpez de Castro et al., 1979), thus establishing the sequence to Trp_{147}. The sequence from Trp_{147} to Glu_{152} was deduced from the sequence of C-11 and T-20 and the amino acid composition of T-19 since T-19 was not obtained in amounts suitable for sequencing. Overlapping peptides (T-20, T-21, C-11, and C-12) extended the sequence to Arg_{170}. Assignment of Arg at position 170 was based on the amino acid composition of T-21 [see Table II in Lôpez de Castro et al. (1979)]. Peptides T-22–T-25 were positioned based on the sequence of peptide

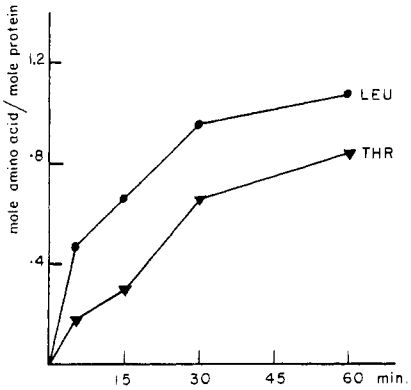

FIGURE 1: Release of amino acids from the HLA-B7$_{pap}$ heavy chain upon digestion with carboxypeptidase A. Only Leu and Thr were released in significant amounts. Other amino acids were detected at less than 1 mol of amino acid per mol of protein after digestion for 60 min.

C-13. Overlap between T-21 and T-22 was provided by the amino-terminal sequence (see Figure 2) of a fragment (I-6) isolated after cleavage at tryptophans with iodosobenzoic acid. The amino-terminal sequence of h3 confirmed the positioning of T-25. Proline-184 and -185 resulted in a high degree of overlap (50% by Val_{189}) during the sequence analysis of h3. This overlap made the assignment of residues during the automatic degradation of h3 difficult and is a likely explanation for the erroneous sequence reported earlier (Terhorst et al., 1977). The sequences of peptides T-26, T-27, C-14, and C-15 confirmed the sequence as deduced from the degradation of h3. In addition, Trägårdh et al. (1979) have recently reported sequence data from this region of HLA and found a sequence consistent with the present one. The positioning of T-28 (residues 220–239) was based on the composition of C-16 (residues 218–241). Although C-16 was isolated in amounts sufficient for sequencing, attempts to obtain its sequence were unsuccessful. A likely explanation may be the cyclization of the amino-terminal Gln_{218} to pyrolidonecarboxylic acid. The remainder of the HLA-B7 sequence, to Lys_{268}, was provided by sequencing overlapping tryptic and chymotryptic peptides.

Since the degradations of T-31 and C-19 failed to reach the carboxyl-terminal residues, the sequence Pro_{269}-Leu_{270}-Thr_{271} at the carboxyl terminus of HLA-B7 was deduced from the compositions of T-31, C-19, and C-20 and carboxypeptidase A digestion. Since C-19 and C-20 had a common amino-terminal sequence, the extra Thr, present in C-20, indicated that Thr is the carboxyl-terminal residue of this peptide. Carboxypeptidase A digestion of HLA-B7 heavy chain released only Leu and Thr (Figure 1). This result agrees with the presence of Pro_{269}, which is not released by carboxypeptidase A (Ambler, 1967). Thus, Thr_{271} is likely the carboxyl terminus of HLA-B7 and Leu_{270} the penultimate residue. While the recovery of two chymotryptic peptides, C-19 and C-20, from the carboxyl terminus of HLA-B7 might be due to partial cleavage at Leu_{270} by chymotrypsin, a more likely explanation is that papain digestion produced a heterogeneous carboxyl terminus. This is supported by the fact that a corresponding tryptic peptide in this region (T-31) does not contain the COOH-terminal threonine as indicated by amino acid composition [see Table III in Lôpez de Castro et al. (1979)].

Peptides Providing Overlaps between Large Fragments. The CNBr and acid cleavages of HLA-B7 heavy chain produced overlapping fragments. Fragment ac-1 provided the overlap for CNBr-1 and CNBr-2 fragments, and CNBr-2 provided the overlap between ac-1 and ac-2 fragments. These fragments were used to obtain the corresponding overlapping peptides.

Table V: Statistical Scores (Z Factors) for HLA-B7 and Human Ig Regions[a]

	α1	α2	α3	β_2m	C	V
α1	–	*5.2*[b]	2.7	1.1	1.5	0.0
α2	*5.2*[b]	–	0.1	0.9	1.1	0.6
α3	2.7	0.1	–	*8.4*	*8.7*	0.8
β_2m	1.1	0.9	*8.4*	–	*5.5*	1.0
C	1.5	1.1	*8.7*	*5.5*	*11.8*	1.6
V	0.0	0.6	0.8	1.0	1.6	*10.5*

[a] α1, α2, and α3 are, respectively, the amino-terminal and first and second disulfide loop regions of the HLA-B7 heavy chain. V, human Ig variable regions [sequences used are V_κI (AG) and V_λIII (SH)]; C, human Ig constant regions [sequences used are C_κ (EU), C_λ (SH), C_γ1, C_γ2, and C_γ3 (EU) and C_μ3 and C_μ4 (GAL)]; all taken from Barker & Dayhoff (1976). Values involving Ig regions are averages of several individual comparisons. Z factors are calculated as described under Materials and Methods. Values of $Z > 3.0$ are underlined to indicate statistical significance. The table is symmetric, each value appearing twice for facile reference. All Z values are obtained with a gap penalty of −2 and a bias parameter of 2. [b] Under stricter gap control parameters, a value of 2.9 is obtained for this pair (see Table VI). In all cases 50 randomizations were performed, except for internal HLA-B7 comparisons where 150 randomizations were performed.

Overlap between CNBr-1 and CNBr-2 was established through the characterization of the tryptic glycopeptide T-13 and its correlation with peptides T-12 and T-14, as discussed in the preceding paper. Overlap between the two acid cleavage fragments (ac-1 and ac-2) was determined by sequencing the chymotryptic peptide, C-13, from CNBr-2. Amino acid composition [Table III in Lôpez de Castro et al. (1979)] suggested that this peptide extended from residue 172 to 201 (Figure 2). The sequence of the first 15 residues confirmed that C-13 started at position 172, spanned the acid cleavage site at Asp_{183}-Pro_{184}, and overlapped with the amino-terminal residues of fragment h3. Therefore, these overlaps establish that, during both CNBr and acid cleavage, no residues are lost due to multiple cleavage at neighboring peptide bonds.

Sequence of the HLA-B7 Heavy Chain. The complete sequence of papain-solubilized HLA-B7 heavy chain is presented in Figure 2, along with the peptides used in obtaining the sequence. This sequence agrees well with the calculated amino acid composition as determined after acid hydrolysis (Lôpez de Castro et al., 1979). The linear arrangement of the two disulfide loops has been previously established (Strominger et al., 1974; Peterson et al., 1975; Terhorst et al., 1977). Therefore, the first disulfide loop of HLA-B7 heavy chain extends from Cys_{101} to Cys_{164} and the second disulfide loop extends from Cys_{203} to Cys_{259}. The distance between loops is 38 residues. The calculated polypeptide molecular weight of 31 333 is in agreement with a size of 33 000 determined for deglycosylated, papain-solubilized HLA-B7 heavy chain by using $NaDodSO_4$ gel electrophoresis (Parham et al., 1977).

Analysis of the HLA-B7 Heavy-Chain Sequence for Homologies. (a) Homology to Other Proteins. The second disulfide loop region of the HLA-B7 heavy chain has been found to be homologous to Ig constant domains (Orr et al., 1979a). This has prompted a search for homology between other regions of the HLA heavy chain and Ig domains. Comparisons have been performed by using the program AL04 (see Materials and Methods) in which the three heavy-chain regions (amino-terminal and first and second disulfide loops) were compared to selected Ig C and V domains. The results are shown in Table V. No homology has been found except the previously documented one between the second disulfide loop region and Ig constant domains and β_2m.

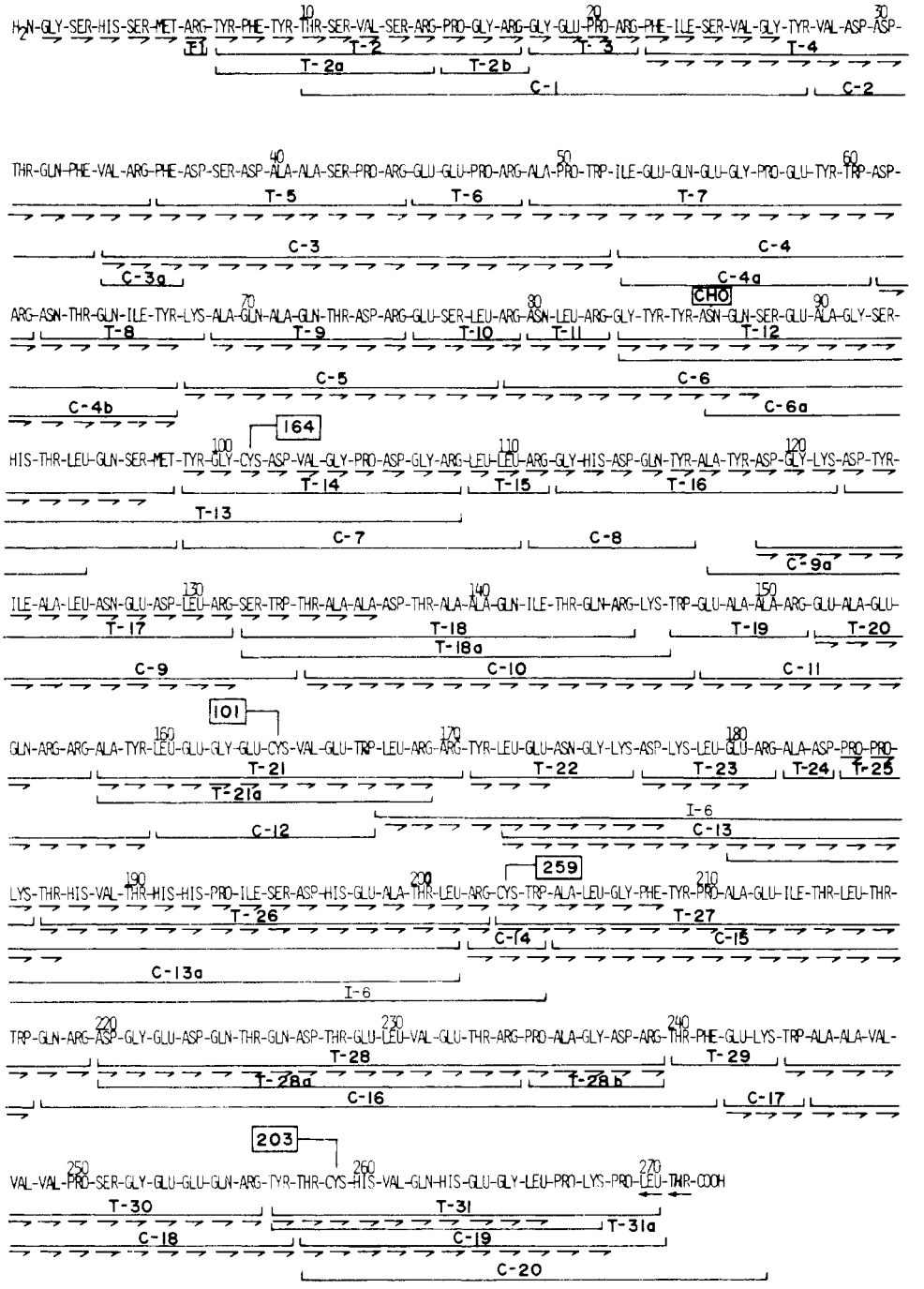

FIGURE 2: Complete sequence of HLA-B7$_{pap}$ heavy chain. Tryptic (T) and chymotryptic (C) peptides used to determine the sequence are indicated. Residues sequenced automatically are indicated by → and those released by carboxypeptidase A by ←. CHO denotes the site of carbohydrate attachment. The disulfide bridges are indicated by Cys □.

The whole Dayhoff protein sequence library was subsequently scanned for homology to 30-residue long segments of the HLA heavy chain (search kindly performed by M. Dayhoff and W. Barker of the National Biomedical Research Foundation, Washington, D.C., using the SEARCH program). The results were in complete agreement with the data of Table V. In addition, no homologies were found with any other sequenced protein.

(*b*) *Internal Homology.* As a first step in the search for internal homologies, program IH04 (see Materials and Methods) was used. Test segments of different length and location in the HLA-B7 heavy chain were aligned without gaps with respect to all possible segments of equal size in the same chain. For test segments starting at position 1 and ranging in length from 50 to 90 residues, the best alignment obtained in this search was when residue 1 aligned with residue 91 (a difference of 90 residues). For a test segment spanning from residues 51 to 90, the best alignment was when positions 51 and 136 matched (a difference of 87 residues). Both alignments could be reconciled by using program AL04, where a single break, including three gaps, was introduced (see Table VI). A Z value of 5.2 was obtained, which is significantly higher than 3.0, considered to be the limit of significance for homology between functionally related proteins (Barker & Dayhoff, 1972).[2] Under stricter gap penalty (see Materials and Methods), no gaps formed in the alignment (Table VI) and

[2] Similarly, when a RELATE-type program (Dayhoff, 1976) is used for assessing internal homology without gap formation, $Z = 3.0$ is taken as the lowest significant value (Barker et al., 1978).

Table VI: Segmental Alignment of HLA-B7 Heavy-Chain Sequence[a]

residues		identical residues	conserved residues	Z
	1 2 3 4 5 6 7 8 9			
1-90	GSHSMRYFYTSVSRPGRGEPRFISVGYVDDTQFVRFDSDAASPREEPRAPWIEQEGPEYWDRNTQIYKAQAQTDRESLRNLRGYYNQSEA			
b				
91-180	GSHTLQSMYGCDVGPDGRLLRGHDQYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQRRAYLEGECVEWLRRYLENGKNKLE	17 (19%)	8	2.9
91-177	GSHTLQSMYGCDVGPDGRLLRGHDQYAYDGKDYIALNEDLRSWTAADTAAQITQRK--WEAAREAEQRRAYLEGECVEWLRRYLENGKN	18 (20%)	14	5.2
91-180	GSHTLQSMYGCDVGPDGRLLRGHDQYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQRRAYLEGECVEWLRRYLENGKNKLE			
182-271	ADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFEKWAAVVVPSGEEQRYTCHVQHEGLPKPLT	8 (9%)	9	0.1
1-90	GSHSMRYFYTSVSRPGRGEPRFISVGYVDDTQFVRFDSDAASPREEPRAPWIEQEGPEYWDRNTQIYKAQAQTDRESLRNLRGYYNQSEA			
182-271	ADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFEKWAAVVVPSGEEQRYTCHVQHEGLPKPLT	5 (6%)	20	2.7

[a] The alignments and Z score calculations were performed by program AL04 as described under Material and Methods. An asterisk indicates identical residues; a dot indicates conserved residues, defined as those related by a mutation data matrix element (Dayhoff et al., 1972; Schwartz & Dayhoff, 1978) higher than 11. The values of these elements correspond to the tendency of residues to replace each other in highly homologous proteins and reflect chemical nature, mutability, and frequency of appearance (Dayhoff et al., 1972). A gap penalty of −2 and a bias parameter of 2 were used throughout. [b] A gap penalty of −4 and a bias parameter of 4 were used.

$Z = 2.9$ was obtained, very close to the limit of 3. Thus, the first and second 90-residue segments of the HLA heavy chain appear to be homologous.

Another prominent alignment in the search by program IH04 was located at the match between residues 1 and 182, suggesting possible homology between the first and the third 90-residue segments. Program AL04 did not introduce gaps in this case, and a Z factor of 2.7 was obtained. This lower Z value, together with the finding of three other alignments with comparable scores (matching residue 1 with residues 52, 77, or 117), renders the significance of this homology more dubious.

In parallel, the segment 1–90 was compared to the whole protein sequence library by using a slightly different mutation data matrix than the one used on our programs (comparison kindly performed by Dr. M. Dayhoff and Dr. W. Barker of the National Biomedical Research Foundation, Washington, D.C., using the SEARCH program). Four HLA segments appeared in the first best 200 comparisons (out of roughly 10^5). These were the 90-residue segments starting at positions 77, 88, 91, and 182. Z factors of 3.2, 3.2, 3.1, and 2.9, respectively, were obtained with relation to the distribution of scores of real proteins. Three out of these four alignments are related either to the suggested homology of the first and second regions (88 and 91) or to the putative relationship of the first and third regions (182).

When the segment 91–180 was searched by program IH04 against the whole sequence, the only prominent alignment score was with residues 1–90, although a local maximum was apparent at the match of residues 91–180 with 182–271. However, an AL04 run confirmed that no significant homology existed between the first and second disulfide loop regions (Table VI).

Discussion

The complete amino acid sequence of the papain-solubilized HLA-B7 heavy chain has been determined. This was achieved by a combination of automatic Edman degradation of large polypeptides, including the whole heavy chain, fragment CNBr-2, and fragment h3, and sequencing of 19 tryptic and 13 chymotrypic peptides out of a total of 33 major tryptic and 22 major chymotryptic peptides. The availability of nanomole amounts of HLA-B7 antigen from a human lymphoblastoid cell line, combined with cleavage methods which generated overlapping large fragments (Terhorst et al., 1977; Lõpez de Castro et al., 1979), has made the HLA-B7 molecule a fortunate choice for sequence studies of histocompatibility antigens. In addition, use of Polybrene (Tarr et al., 1978; Klapper et al., 1978) enabled nanomole amounts of peptides to be sequenced, in most cases, to their carboxyl-terminal residues.

The papain-solubilized HLA-B7 heavy chain is 271 residues long. Papain cleaves the molecule at positions 270 and 271, resulting in a heterogeneous carboxyl terminus. It is not known whether other peptide bonds are also cleaved by the enzyme between position 271 and the cluster of hydrophobic residues in the membrane (Springer & Strominger, 1976). A carboxyl-terminal serine residue, as reported previously (Henriksen et al., 1976), was not found in the present study during carboxypeptidase digestion or in the characterization of peptides corresponding to the carboxyl-terminal part of the molecule. It is possible that slight differences in the carboxyl terminus may be obtained, depending upon specific papainolysis conditions.

On the basis of the general features of the polypeptide chain, the papain-solubilized HLA-B7 heavy chains can be considered

in terms of three regions of ~90 residues each: (1) the amino-terminal region, with no disulfide bonds, which extends approximately to the single, N-linked, carbohydrate moiety at Asn_{86}; (2) the region of the first disulfide loop; (3) the region of the second disulfide loop. Fortuitously, the existence of a methionine residue at position 98 and of an acid-labile aspartyl–proline bond at positions 183–184 permitted the chemical dissection of these three regions.

The disulfide bonds span residues 101–164 and 203–259, thus defining two loops of 62 and 55 residues, respectively, which are 38 residues apart. The analogy of this situation with the immunoglobulin molecule was noted earlier and prompted the speculation that the HLA molecule may also consist of several domains (Strominger et al., 1974; Peterson et al., 1975). For investigation of the possibility of internal homology regions, the HLA-B7 heavy-chain sequence was searched by using computer methods designed to detect distant evolutionary relationships (Dayhoff et al., 1976). This analysis resulted in the detection of possible homology between the amino-terminal region (residues 1–90) and the region of the first disulfide bond (residues 91–180). The statistical significance of this homology was evaluated in terms of the Z factor and compared with the minimum score of $Z = 3.0$ SD which may be taken as indicative of internal duplication (Barker & Dayhoff, 1972; Barker et al., 1978). On the basis of this criterion, the value of $Z = 2.9$, obtained for the alignment of these two segments without gaps, is suggestive of internal homology. The fact that the introduction of only a single gap of three residues in the alignment significantly improves the alignment score strongly suggests that the amino-terminal portion and the region of the first disulfide loop may constitute two homology regions of the molecule.

It is interesting that, in the alignment of the first and second heavy-chain regions (with gaps introduced), the two cysteines of the second region match the two serines in the first, this substitution being one of the most favorable for cysteines by the mutation data matrix (Schwartz & Dayhoff, 1978). Also, in H-2 antigens, the murine counterparts of HLA, the two carbohydrate moieties appear in what would be homologous positions of the first and second heavy-chain regions by our alignments (Nathenson et al., 1979). It should, however, be pointed out that, although the above structural implications are likely to be general, other sequences of HLA or its counterparts in other species may display different degrees of internal homology, since both the first and the second heavy-chain regions contain some variability (Orr et al., 1979b).

When residues 1–90 were compared with the region of the second disulfide loop (residues 182–271), a Z score of 2.7 SD was obtained. No improvement was obtained by a controlled allowance for gaps. Therefore, the significance of this homology appears dubious. Similar analysis fails to detect any homology between the first and the second disulfide loop regions. It is possible that these two regions may have analogies at the three-dimensional level in spite of a lack of detectable sequence homology, as is the case, for instance, between immunoglobulin V and C regions (Poljak et al., 1973) and between the chymotrypsin domains (McLachlan, 1979).

A picture emerges in which the papain-solubilized HLA-B7 antigen may be composed of four domains, three in the heavy chain and one in the light chain. This situation is somewhat different from the earlier suggestion of two extracellular heavy-chain domains (Strominger et al., 1974; Peterson et al., 1975). Two of these domains, β_2m and the second disulfide loop region of the heavy chain, are homologous to each other, as well as to Ig constant domains (Strominger et al., 1978; Orr et al., 1979a). The two others (amino-terminal and first disulfide loop regions) are also homologus to each other, but not to Ig domains or to the other regions of the molecule. It is conceivable, based on the organization of the Ig molecule [see, for example, Davies et al. (1975)], that homologous domains in the HLA molecule may also interact extensively. However, the isolation (although in only 20% yield) of a fragment containing β_2m, but lacking the second disulfide loop region (Trägårdh et al., 1979), suggests that β_2m interacts significantly with nonhomologous regions of the molecule. A difference in the quaternary interactions between HLA and Ig is also suggested by the fact that the conformation of the heavy chain appears to be very dependent upon interactions with β_2m (Ploegh et al., 1979; Lancet et al., 1979; Krangel et al., 1979).

The primary structure of the HLA-B7 heavy chain does not provide incontestable data to formulate a mechanism for the evolution of the histocompatibility antigens. However, on the basis of the detectable homology between the amino-terminal and the first disulfide loop regions, it would appear plausible that a distant gene duplication of an ancestral region coding for ~90 amino acid residues took place during the evolution of the molecule. The same mechanism may be invoked in relation to the second disulfide loop region, as suggested by the similar size of both disulfide loops.

If this is the case, a common ancestry between histocompatibility antigens and immunoglobulins could be postulated, given the convincing evidence of a structural relationship between the second disulfide loop region and immunoglobulin constant domains and β_2-microglobulin (Strominger et al., 1978; Orr et al., 1979a). However, the fact that the first two regions of HLA-B7 heavy chain are homologous neither to the second disulfide loop region and β_2m nor to Ig constant or variable domains requires that this hypothesis be considered with caution. The recent elucidation of the importance of gene splicing in the Ig system (Sakano et al., 1979) offers an additional mechanism of possible relevance in the evolution of HLA antigens.

A comparison of the sequence of HLA-B7 heavy chain with a partial sequence of the murine H-$2K^b$ heavy chain (Coligan et al., 1979; see Table VII) reveals a high degree of conservation in the primary structure: 31 differences out of 96 residues compared (68% homology). The substitutions are scattered throughout the polypeptide chain. However, a striking cluster of differences appears between residues 60 and 82, where 14 substitutions are present in a segment of 22 residues (36% homology). Large differences in amino acid sequence are observed with respect to the same region in HLA-A2 heavy chain (Orr et al., 1979b). It is possible that this region may contribute significantly to the alloantigenic determinants of the molecule. Obviously, many more sequences are necessary to define precisely the molecular nature of the alloantigenic site(s), but this clustering of differences certainly suggests that the allotypic substitutions may lie in a limited number of positions in the heavy chain which are at the surface of the native molecule. In this sense, it may be relevant to note that the attachment point for the carbohydrate moiety of the molecule, Asn_{86}, is close to the segment in which unusual variability is observed. It is conceivable that the three-dimensional structure of the histocompatibility antigen molecule provides for a rigid framework capable of accommodating a large number of amino acid substitutions in regions of the surface, without significant variations in the overall conformation.

Table VII: Comparison of Partial Sequences[a] of HLA-B7 and H-2K[b]

	1	2	3	4	5	6	7	8	9
HLA-B7	GSHSMRYFYTSVSRPGRGEPRFISVGYVDDTQFVRFDSDAASPREEPRAPWIEQEGPEYWDRNYQIYKAQAQTDRESLRNLRGYYNQSEAGSHTLQ								
H-2K[b]	-P-L---V-A----L----YME----------E----------EN---Y----R-M----------E-ET-KA-GNE-SF-VD--T-L-------KG----I-								

[a] H-2K[b] sequence data are from Coligan et al. (1979).

References

Ambler, R. (1967) *Methods Enzymol. 11*, 436–445.

Barker, W. C., & Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, pp 101–110, National Biomedical Research Foundation, Washington, D.C.

Barker, W. C., & Dayhoff, M. O. (1976) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 2, pp 165–190, National Biomedical Research Foundation, Washington, D.C.

Barker, W. C., Ketchman, L. K., & Dayhoff, M. O. (1978) *J. Mol. Evol. 10*, 265–281.

Brauer, A. W., Margolies, M. N., & Haber, E. (1975) *Biochemistry 14*, 3029–3035.

Coligan, J. E., Kindt, T. J., Ewenstein, B. M., Uehara, H., Martinko, J. M., & Nathenson, S. G. (1979) *J. Mol. Immunol. 16*, 3–8.

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. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 2123–2127.

Davies, D. R., Padlan, E. A., & Segal, D. M. (1975) *Contemp. Top. Mol. Immunol. 4*, 127–155.

Dayhoff, M. O. (1976) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 2, pp 1–8, National Biomedical Research Foundation, Washington, D.C.

Dayhoff, M. O., Eck, R. V., & Park, C. M. (1972) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, pp 89–99, National Biomedical Research Foundation, Washington, D.C.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1976) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 2, pp 9–19, National Biomedical Research Foundation, Washington, D.C.

Ferguson, W. S., Terhorst, C. T., Robb, R. J., & Strominger, J. L. (1979) *Mol. Immunol. 16*, 23–28.

Grey, H. M., Kubo, R. T., Colon, S. M., Poulik, M. D., Cresswell, P., Springer, T., Turner, M., & Strominger, J. L. (1973) *J. Exp. Med. 138*, 1608–1612.

Henriksen, O., Apella, E., Smith, D. F., Tanagaki, N., & Pressman, D. (1976) *J. Biol. Chem. 251*, 4214–4219.

Klapper, D. G., Wilde, C. E., III, & Capra, J. D. (1978) *Anal. Biochem. 85*, 126–131.

Krangel, M. S., Orr, H. T., & Strominger, J. L. (1979) *Cell* (in press).

Kronquist, K. K., & Lennarz, W. J. (1978) *J. Supramol. Struct. 8*, 51–65.

Kulbe, K. D. (1974) *Anal. Biochem. 59*, 564–573.

Lancet, D., Parham, P., & Strominger, J. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3844–3848.

Lôpez de Castro, J. A., Orr, H. T., Robb, R. J., Kostyk, T. G., Mann, D. L., & Strominger, J. L. (1979) *Biochemistry* (preceding paper in this issue).

Mahoney, W. C., & Hermodson, M. A. (1979) *Biochemistry 18*, 3810–3814.

McLachlan, A. D. (1979) *J. Mol. Biol. 128*, 49–79.

Nakamuro, K., Tanigaki, N., & Pressman, D. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 2863–2865.

Nathenson, S. G., Ewenstein, B. M., Uehara, H., Martinko, J. M., Coligan, J. E., & Kindt, T. J. (1979) in *Trends in Histocompatibility* (Reisfeld, R. A., & Ferrone, S. O., Eds.) (in press).

Needleman, S. B., & Wunsch, C. D. (1970) *J. Mol. Biol. 43*, 443–453.
Niall, H. D. (1973) *Methods Enzymol. 27*, 942–1010.
Orr, H. T., Lancet, D., Robb, R. J., Lôpez de Castro, J. A., & Strominger, J. L. (1979a) *Nature (London)* (in press).
Orr, H. T., Lôpez de Castro, J. A., Parham, P., Ploegh, H. L., & Strominger, J. L. (1979b) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4395–4399.
Parham, P., Alpert, B. N., Orr, H. T., & Strominger, J. L. (1977) *J. Biol. Chem. 252*, 7555–7567.
Peterson, P. A., Cunningham, B. A., Berggard, I., & Edelman, G. M. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1697–1701.
Peterson, P. A., Rask, L., & Lindblom, J. B. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 35–39.
Peterson, P. A., Rask, L., Sege, K., Klareskog, L., Anundi, H., & Ostberg, L. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1612–1616.
Pisano, J. J., & Bronzert, T. J. (1969) *J. Biol. Chem. 244*, 5597–5607.
Ploegh, H. L., Cannon, L. E., & Strominger, J. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2273–2277.
Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P., & Saul, F. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3305–3310.
Sakano, H., Rogers, J. H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R., & Tonegawa, S. (1979) *Nature (London) 277*, 627–633.
Sauer, R. T., Niall, H. D., Hogan, M. L., Kentman, H. T., O'Riordan, J. L. H., & Potts, J. T., Jr. (1974) *Biochemistry 13*, 1994–1999.
Schwartz, R. M., & Dayhoff, M. O. (1978) in *Origin of Life* (Noda, H., Ed.) pp 457–469, Japan Science Society Press, Tokyo.
Smithies, O., & Poulik, M. D. (1972) *Science 175*, 187–189.
Smithies, O., Gibson, D., Fanning, E., Goodfliesh, R., Gilman, J., & Ballantyne, D. (1971) *Biochemistry 10*, 4912–4921.
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.
Strominger, J. L., Kabat, E. A., Bilofsky, H., Orr, H., Parham, P., Ploegh, H., Robb, R., Terhorst, C., & Wu, T. T. (1978) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 37*, 1467.
Summers, M. R., Symthers, G. W., & Oroszlan, S. (1973) *Anal. Biochem. 53*, 624–628.
Tarr, G. E., Beecher, J. F., Bell, M., & Mckean, D. J. (1978) *Anal. Biochem. 84*, 622–627.
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.
Trägårdh, L., Wiman, K., Rask, L., & Peterson, P. A. (1979) *Biochemistry 18*, 1322–1328.

Cell-Specific Antigens in Chicken Erythroid Nuclei: Species Specificity†

Wanda M. Krajewska, Robert C. Briggs, and Lubomir S. Hnilica*

ABSTRACT: Antisera raised to dehistonized chicken reticulocyte chromatin were tested for their cell and species specificity. Quantitative microcomplement fixation and immunohistochemical localization revealed the presence in chromatin of erythroid cell-specific nonhistone protein antigen(s). The antigenic specificity was shown to depend on the association of the antigenic protein(s) with deoxyribonucleic acid (DNA). Although the antisera were exceptionally cell specific, they cross-reacted with erythroid cells of other avian species. The extent of cross-reactivity was found to approximate the phylogenetic distances of the tested avian species. Erythroid cells from fish and amphibians were not reactive. Reconstitution experiments of partially purified chicken reticulocyte chromosomal nonhistone protein antigens with DNAs isolated from several vertebrate species showed that the species specificity of the antigenic complexes is determined principally by the species origin of the nonhistone proteins. Our results show that a cell-specific chromosomal nonhistone protein(s) has undergone evolutionary change and the relative immunological differences observed are consistent with the accepted phylogenetic distances of the species examined.

Production of antibodies against cellular antigens represents one of the most powerful tools for probing macromolecular diversity and specificity. While most of the tissue antigens described in the literature are of extracellular or cytoplasmic origin, the presence of specific antigens in the cell nucleus has only recently been appreciated.

First notions that chromatin of eucaryotic cells may contain cell- or tissue-specific antigens were pointed out by Henning et al. (1962) and Messineo (1961), who studied the immunological properties of nucleohistones isolated respectively from calf thymus and leukocyte nuclei. Development of techniques for isolation of chromatin permitted detailed studies on its biological and structural properties. Using dehistonized chromatin from chick oviduct, Chytil & Spelsberg (1971) elicited antibodies which specifically recognized chromatin from this tissue. The immunological specificity of chromosomal complexes was confirmed by Wakabayashi & Hnilica (1973), who found that maligant growth significantly altered the antigenic properties of rat liver chromatin. A similar observation was also made by Zardi et al. (1973), who reported the antigenic properties of chromatin from SV40-transformed WI 38 cells to be different from those of the nontransformed controls. According to Wakabayashi et al. (1974), the anti-

† From the Department of Biochemistry and the A. B. Hancock, Jr., Memorial Laboratory, Vanderbilt University School of Medicine, Nashville, Tennessee 37232. *Received May 8, 1979.* This work was supported by National Cancer Institute Grant CA-18389.

0006-2960/79/0418-5720$01.00/0