Proc. Natl. Acad. Sci. USA
Vol. 76, No. 9, pp. 4395–4399, September 1979
Biochemistry

Comparison of amino acid sequences of two human histocompatibility antigens, HLA-A2 and HLA-B7: Location of putative alloantigenic sites

(protein structure/genetic polymorphism/alloantigenic determinants)

HARRY T. ORR, JOSE A. LOPEZ DE CASTRO, PETER PARHAM, HIDDE L. PLOEGH, AND JACK L. STROMINGER

The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Contributed by Jack Leonard Strominger, June 27, 1979

ABSTRACT **The complete amino acid sequence for papain-solubilized HLA-B7 heavy chain is compared with the partial sequences of HLA-A2 and H-2K^b heavy chains. Although these molecules are highly conserved (i.e., 80% homology in comparing HLA-B7 with HLA-A2; 72% and 74% homology in comparing H-2K^b with HLA-A2 and HLA-B7, respectively), two stretches of greater variability are observed. These clusters of variability are discussed in terms of their possible involvement in the alloantigenic determinant(s) characteristic of these highly polymorphic membrane antigens.**

The major histocompatibility complex of humans and mice encodes several membrane antigens that play a central role in the immune response (1). These include the protein products of the *HLA-A* and *HLA-B* loci in humans and the *H-2K* and *H-2D* loci in mice. The most striking feature of this system is its genetic polymorphism, which gives rise to over 20 alleles at each of the *HLA-A* and *HLA-B* loci and to much larger numbers at the *H-2K* and *H-2D* loci. This polymorphism is believed to be related in some manner to the principal functions of these molecules, including roles in cell–cell interactions and as participants in cell-bound defense mechanisms. Presumably the evolutionary pressure for the polymorphism was provided by a variety of environmental stresses. These might include the need to control the proliferation of modified syngeneic cells (e.g., virus-infected or tumor cells) and of semi-allogeneic cells arising from the highly invasive trophoblast in placental mammals.

Our understanding of the biochemical nature of these molecules has progressed substantially in the last few years (2). They are glycoproteins consisting of two polypeptide subunits (3). A large 44,000-dalton glycosylated polypeptide is associated with a nonglycosylated 12,000-dalton polypeptide. The smaller subunit is invariant, is not encoded by the major histocompatibility complex, and is identical to β_2-microglobulin (4–7). On the other hand, the large subunit is encoded by the major histocompatibility complex and contains the alloantigenic determinants characteristic of these antigens. Previous experiments have shown that, in both mice (8) and humans (9), these determinants are not located in the carbohydrate portion of the heavy chain and thus must reside in the polypeptide chain. One of the most interesting features of the sturcture of these molecules is the location of these alloantigenic site(s) in the polypeptide, because they might be the functional regions of the molecules that have varied during evolution. They may have some recognition or receptor function (just as the hypervariable regions of Ig are the sites of antigen binding).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Papain enzymatically releases from the membrane a soluble complex, HLA_{pap}, in which the heavy chain is 34,000 daltons and the small subunit, β_2-microglobulin, is uncleaved and in which the alloantigenic activity is unaltered (10, 11). This method of solubilization combined with the use of the human lymphoblastoid cell line JY has provided material for the elucidation of the complete amino acid sequence of the $HLA\text{-}B7_{pap}$ heavy chain (271 residues). A second antigen, HLA-A2, is also easily obtained from this cell line. A virtually complete amino acid sequence of the amino-terminal 167 residues of the HLA-A2 heavy chain has also been obtained. In order to determine the possible location of the variable alloantigenic determinant(s), these sequences are compared to each other and to data presently available for the H-2K^b antigen (12).

MATERIALS AND METHODS

Cell Line. The human lymphoblastoid cell line JY, homozygous at both the *HLA-A* and the *HLA-B* loci (A2,2;B7,7), was used.

Preparation of Papain-Solubilized HLA Antigens (HLA_{pap}). HLA antigens were solubilized from JY membranes and purified as described (9). The two subunits of HLA were separated by using a Sephadex G-75 column in 1 M acetic acid.

Cyanogen Bromide Cleavage. The $HLA\text{-}A2_{pap}$ heavy chain was cleaved with CNBr in 70% (vol/vol) formic acid for 16 hr at 25°C either before or after reduction and alkylation of cysteines (9). When cleavage preceded reduction and alkylation, the CNBr fragments were applied directly to a Sephadex G-100 column and eluted with 1 M propionic acid. In the case in which cleavage followed reduction and alkylation, the CNBr fragments were lyophilized, redissolved in 200 mM NH_4HCO_3, and spun at 1000 × *g*, and the supernatant was applied to a Sephadex G-75 superfine column.

Enzymatic Generation of Peptides. Cyanogen bromide fragments were digested with trypsin (enzyme:substrate = 1:33 wt/wt, 6 hr, 37°C) and chymotrypsin (enzyme:substrate = 1:200 wt/wt, 30 min, 25°C). These peptides were isolated by using a microbore ion-exchange column eluted with a pyridine/acetic acid gradient (13). Tryptic peptides were also obtained from the CNBr-3 fragment after blockage of arginine residues with cyclohexanedione (14). These peptides were separated on a column of Sephadex G-50 in 1 M acetic acid. The use of fluorescamine in the detection of peptides was as described (15).

Amino Acid Analysis. The compositions of fragments and peptides were determined on a Beckman 121M automatic amino acid analyzer after hydrolysis for 24 hr at 110°C with 5.7 M HCl and 0.1% 2-mercaptoethanol.

Abbreviation: HLA_{pap}, papain-solubilized HLA antigen.

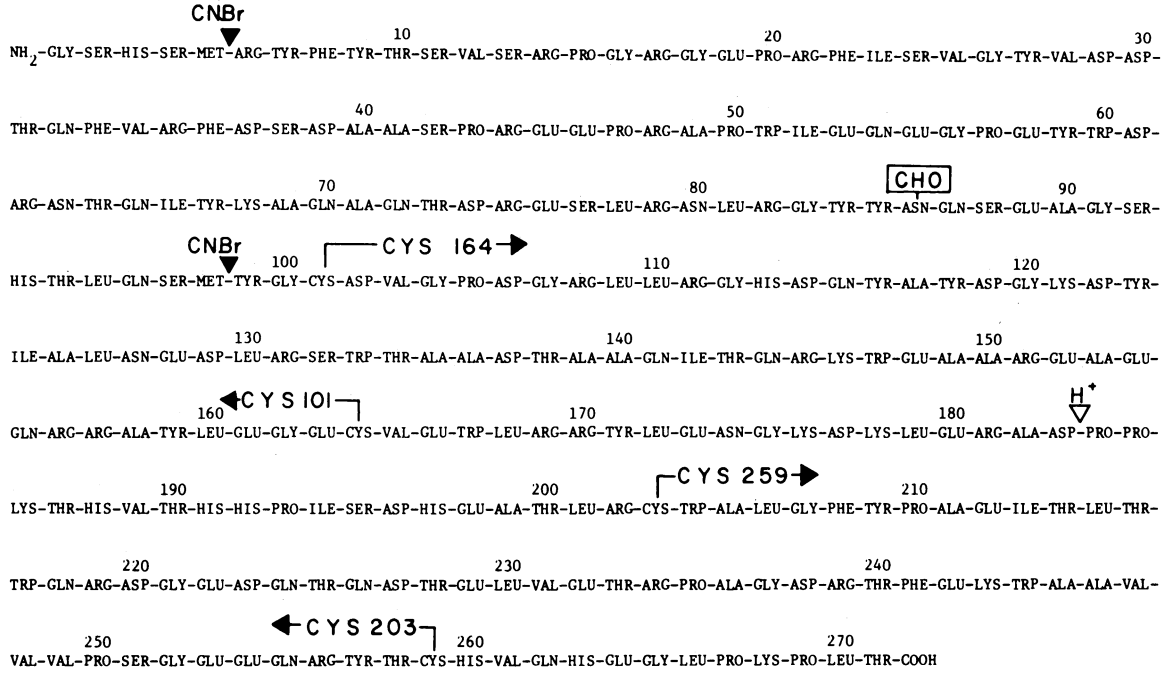

FIG. 1. The complete amino acid sequence of HLA-B7$_{pap}$. Cyanogen bromide and acid (H^+) cleavage points (26) are indicated. Also shown are the positions of the carbohydrate moiety (CHO) and the arrangement of the disulfide loops.

Sequence Determination. Automatic degradations of large fragments were performed by using an updated Beckman 890B sequenator. The 0.1 M Quadrol program (16) was utilized and degradations were performed after the addition of 3 mg of Polybrene to the cup (17, 18). Two methods were used to perform manual degradations of peptides: (*i*) the three-step manual Edman method as modified by Sauer *et al.* (19) and (*ii*) the dansyl-monitored Edman method (20). In the case of automatic degradations, phenylthiohydantoin residues were identified by using thin-layer chromatography (21, 22), gas/liquid chromatography (23, 24), and back hydrolysis followed by amino acid analysis (25).

RESULTS AND DISCUSSION

Complete Amino Acid Sequence of the Papain-Solubilized Heavy Chain of HLA-B7. The complete sequence of this molecule is shown in Fig. 1. The sequence was obtained from tryptic and chymotryptic peptides of the three large fragments derived from the heavy chain of HLA-B7 by acid and CNBr cleavages (26). A full account of this sequence determination and its analysis will be presented elsewhere.*

Cyanogen Bromide Cleavage of HLA-A2$_{pap}$ Heavy Chain. Unlike HLA-B7 and several other HLA antigens, the heavy chain of HLA-A2 does not contain an acid-labile Asp-Pro bond (26). Therefore, CNBr was chosen for the generation of large fragments for sequencing HLA-A2. Amino acid analysis of HLA-A2 heavy chain indicated the presence of four methionines (27), which should result in five CNBr fragments. Fig. 2 presents the elution profile of CNBr fragments obtained from HLA-A2 heavy chain without prior reduction and alkylation. Pool II was subsequently found to contain two CNBr fragments linked by a disulfide bridge. This was established by the presence of two NH_2-terminal residues, Tyr and Ala, in this pool. In addition, CNBr cleavage after reduction and alkylation yielded fragments CNBr-4 and CNBr-5, which were separated by gel filtration. Analysis of the remaining pools (Fig. 2) showed that each contained a single CNBr fragment, with pool III containing CNBr-3, pool IV CNBr-2, and the included peak, pool V, containing CNBr-1.

Alignment of CNBr fragments 1 and 2 was accomplished from the amino-terminal sequence of HLA-A2 (27). The isolation and sequence of the tryptic peptide Met-Glx-Pro-Arg from HLA-A2 heavy chain provided the overlap between CNBr-2 and CNBr-3. The remaining CNBr fragments, 4 and 5, were aligned on the basis of comparison with the HLA-B7 sequence.

Sequence Analysis of the Amino-Terminal 167 Residues of the HLA-A2 Heavy Chain. The sequence of the amino-terminal 167 residues of HLA-A2$_{pap}$ heavy chain is shown in Fig. 3. Edman degradation of HLA-A2 heavy chain combined

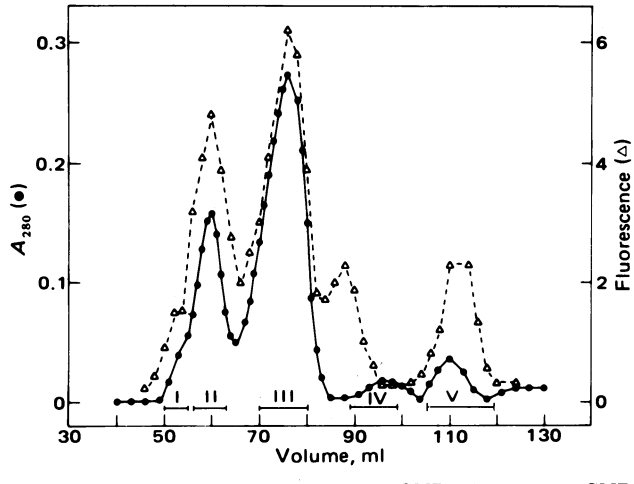

FIG. 2. Purification of HLA-A2$_{pap}$ CNBr fragments. CNBr cleavage and fragment isolation were performed without reduction and alkylation. Chromatography was on a 1 × 200 cm column of Sephadex G-100 with 1 M propionic acid. Fragments were located by A_{280} and by fluorescence (scale is arbitrary) after reaction with fluorescamine. Horizontal bars indicate pooled fractions.

* H. T. Orr, J. A. Lopez de Castro, D. Lancet, and J. L. Strominger, unpublished.

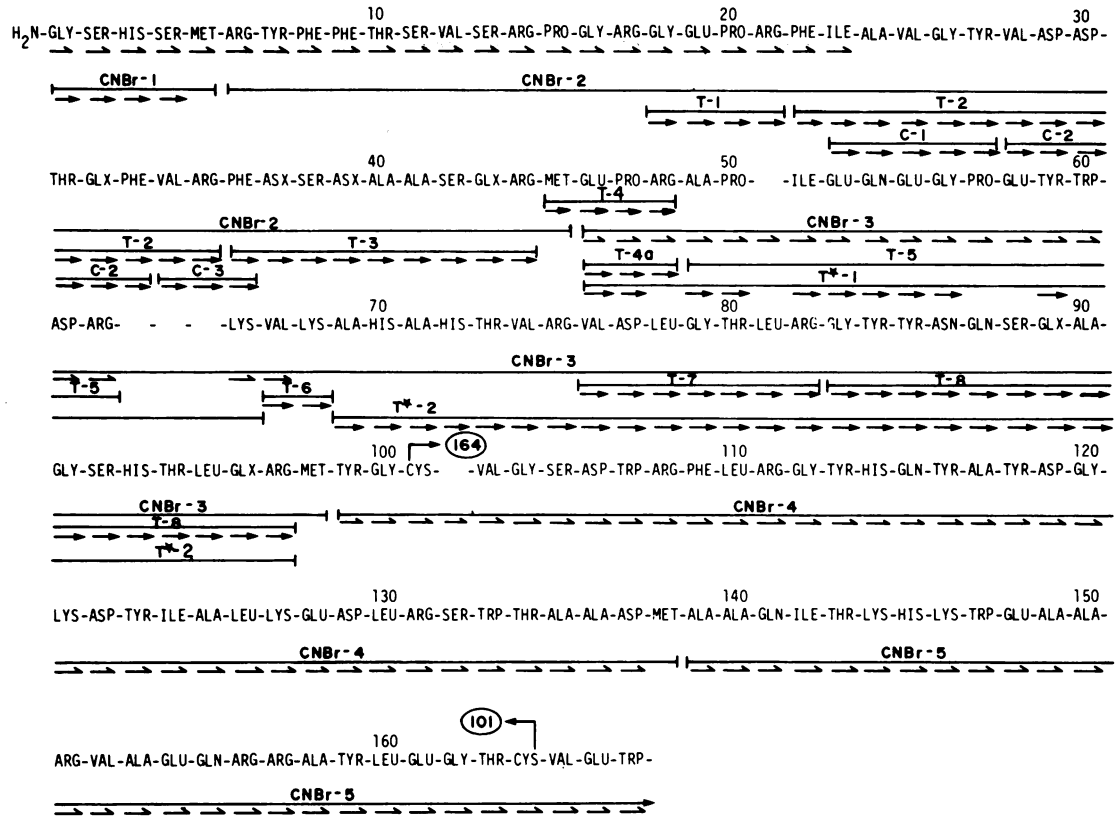

FIG. 3. Sequence of HLA-A2$_{pap}$ to residue 167. T denotes tryptic peptides; T*, tryptic peptides generated after Arg blockage; and C, chymotryptic peptides. Positions sequenced automatically are indicated by (—) and manually by (→).

with overlapping tryptic, chymotryptic, and CNBr peptides provided sequence data to residue 98. The sequence from 99 to 138 was determined from the automatic degradation of CNBr-4 to its carboxy terminus. Residues 139–167 were determined by automatic Edman degradation of CNBr-5, which by amino acid composition and comparison with HLA-B7, is thought to extend to residue 271 (unpublished data). The absence of homoserine and homoserine lactone after acid hydrolysis of CNBr-5 also indicated that this fragment extends to the carboxy terminus of HLA-A2$_{pap}$ heavy chain. The sequence of HLA-B7 heavy chain was used to align these two sequences, 99–138 and 139–167. At five positions no assignment of amino acid residues has been made yet. Therefore, the data provide the sequence of the amino-terminal two-thirds of the HLA-A2$_{pap}$ heavy chain with the exception of these residues.

The amino-terminal region of the HLA-A2 heavy chain, residues 1–98, contains the single Asn-linked carbohydrate of HLA-A2 (9). Like HLA-B7, this carbohydrate is attached to Asn-86. The tripeptide sequence Asn-Gln-Ser at residues 86–88 of HLA-A2 is compatible with a sequence of Asn-X-Ser/Thr noted for other glycoproteins at the site of glycosylation (28). Residues 99–167 contain the first disulfide loop of the HLA-A2 heavy chain. The linear arrangement of the disulfide loops in HLA-A2 is based on the dissociation of CNBr-4 and CNBr-5, each containing a single Cys residue, after reduction and alkylation. Thus, as in HLA-B7, the first disulfide loop of HLA-A2 heavy chain extends from Cys-101 to Cys-164, forming a loop of 63 residues.

The finding of significant sequence homology between the second disulfide loop of HLA-B7 heavy chain and immunoglobulin (Ig) constant domains (13) prompted a preliminary search for such homology with the HLA-A2 sequence. As with the HLA-B7 sequence, no homology between either Ig constant or variable domains and the amino-terminal 90 residues of HLA-A2 heavy chain was detected. Similarly, the first disulfide loop of HLA-B7 also failed to show Ig homologies.*

Comparison of HLA-A2$_{pap}$ Heavy Chain Data with Those from HLA-B7$_{pap}$ and H-2K^b Antigens. As discussed earlier (8, 9), it is clear that the polypeptide chain of human and mouse histocompatibility antigens contain the alloantigenic determinants. The extensive sequence data available for two HLA antigens and an H-2 antigen afford the opportunity to assess the nature of primary structural differences that may contribute to the alloantigenic determinants.

Fig. 4 compares the complete sequence of HLA-B7$_{pap}$ heavy chain with the available HLA-A2 and H-2K^b (12) sequences. In comparing the first 167 residues in HLA-B7 with those in HLA-A2, an overall sequence homology of 80% is obtained. Comparison of the available residues in H-2K^b with those of HLA-B7 and HLA-A2 results in sequence homologies of 74% and 72%, respectively. Thus, the level of sequence homology between H-2 and HLA is only slightly less than that seen between HLA-A and HLA-B. A closer examination shows that H-2K is no more like HLA-A than it is like HLA-B—i.e., there are six positions at which H-2K^b is identical to HLA-A2 and not HLA-B7 (residues 66, 76, 80, 138, 144, 163) and seven other positions at which H-2K^b is like HLA-B7 and not HLA-A2 (residues 43, 70, 72, 77, 79, 145, 152). Because these residues are intermixed, there is neither an HLA-A region nor an HLA-B region in the H-2K^b sequence. The high level of sequence homology between H-2 and HLA antigens, assessed from comparisons using sequence data from both disulfide loops and the amino-terminal region, clearly show that the human and mouse histocompatibility antigens must be functionally homologous. Such sequence homology also strongly supports the conclusion that an alloantigenic site(s) as it exists in H-2 is likely to be very similar to that in HLA.

The distribution of sequence differences found in comparing

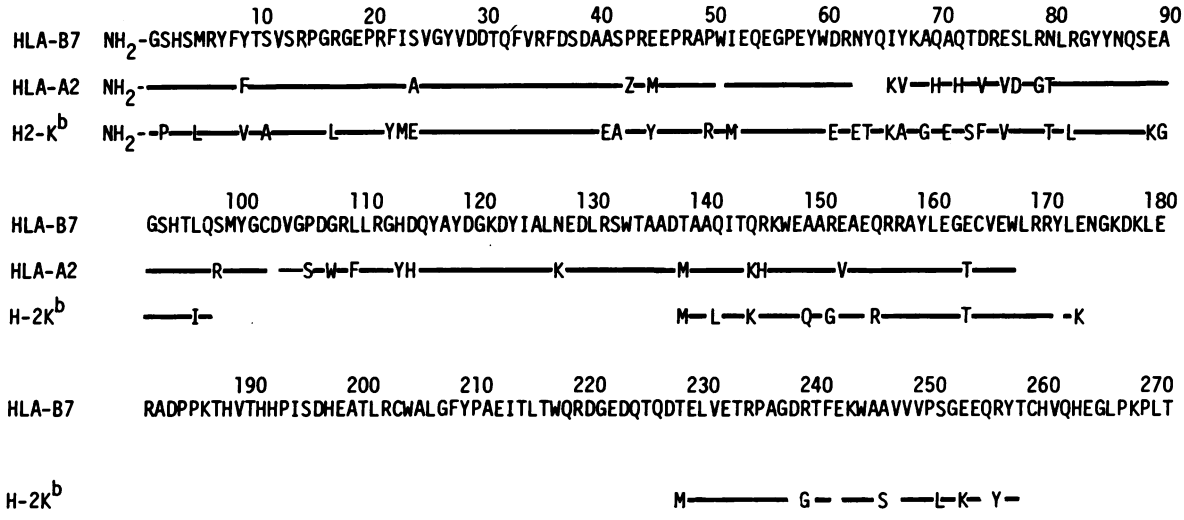

FIG. 4. Comparison of the HLA-B7$_{pap}$ sequence with data on HLA-A2$_{pap}$ and H-2K^b (12). Identical residues are indicated by the solid bars. Residues are in the single-letter code: A, Ala; B, Asx; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; and Z, Glx.

the two HLA specificities presents an interesting pattern. While differences are scattered throughout the polypeptide chains, there is a clear pattern of highly homologous regions spaced between regions of greater variability (Fig. 4). The amino-terminal 60 residues are a highly conserved stretch of the polypeptide chain. There are only 4/60 positions at which HLA-B7 differs from HLA-A2 in this region, yeilding a sequence homology of 93%. This region is followed by a cluster of differences located between residues 60 and 80. In this stretch, there are 9 positions at which HLA-B7 and HLA-A2 differ. These differences are all located within a 15-residue stretch, resulting in a drop in homology to 40%. Only one difference exists between HLA-B7 and HLA-A2 after residue 80 until residue 105, where a second cluster of differences is found. Between residues 105 and 114, 5 out of 10 positions differ. The sequence homology for this region is 50%. In the final stretch of available sequence data, residues 115 to 167, the homology is again at a high level, 88%. Of the 24 residues that differ between HLA-B7$_{pap}$ and the portion of HLA-A2$_{pap}$ whose sequence has been determined, 14 residues (58%) are found in the two clusters. It is also interesting to note that these two clusters are of similar size and have comparable levels of sequence homology.

A similar cluster of variability is found in comparing the HLA-B7 and HLA-A2 sequences with H-2K^b (12). Between residues 60 and 80 of HLA-B7 and H-2K^b heavy chains there are 11 differences. In comparing HLA-A2 with H-2K^b, this same stretch contains 10 differences. Therefore, it is highly possible that the alloantigenic site(s) consists of clusters of residues linearly located in the primary structure. Comparison of HLA with H-2 antigens indicates that one such cluster might be between residues 60 and 80. The HLA-A2 and HLA-B7 data indicate that more than one cluster might contribute to an alloantigenic specificity.

It is likely that the second disulfide loop region of these molecules is more highly conserved than the other regions. This region of HLA-B7 heavy chain shows highly significant sequence homology to Ig constant domains (13), and preliminary data for HLA-A28 and -B40 indicate a high degree of conservation in the second-disulfide loop region. Thus, it is unlikely that any additional large differences will be found between HLA-A2 and HLA-B7 in this region. Thus, the amino-terminal region of HLA antigens and the first disulfide loop region—i.e., residues 1–180—are the likely locations of the alloantigenic residues. The two clusters of variability shown in Fig. 4 are the most striking features seen in comparing HLA-A2 and HLA-B7.

A word of caution must be added regarding the variability of these molecules due to the fact that gene products of different loci are being compared. Therefore, isotypic differences may be superimposed on allotypic differences. In addition, several properties of HLA-A2 make it unique—i.e., lack of an acid cleavage site (26) and its behavior in ion-exchange chromatography (11). Such properties would be reflected in its amino acid sequence. It is clear that a full definition of the molecular basis of the alloantigenic site(s), the extent to which the clusters of variability participate in these determinants, and the differentiation between allotypic and isotypic markers of histocompatibility antigens must await additional evidence. Determining the sequences of the specificities HLA-A28 and HLA-B40 should be particularly useful in this respect, because they are crossreactive with HLA-A2 and HLA-B7, respectively (29). Thus, the number of sequence differences between these crossreactive specificities should be more restricted, perhaps enabling a clearer localization of the alloantigenic residues. In addition, the recent development of HLA-A2 (30) and HLA-B7 (31) specific monoclonal antibodies will provide access to extremely useful technology for the study of alloantigenic determinants.

We thank Mr. Thomas Kostyk for the isolation of the HLA antigens. This work was supported by National Institutes of Health Grant AI 10736, U.S. Public Health Service Research Fellowship AI 05483 to H.T.O., and Public Health Service International Research Fellowship TW 02582 to J.A.L.C.

1. Paul, W. E. & Benacerraf, B. (1977) *Science* **195**, 1293–1300.
2. Strominger, J. L., Engelhard, V. H., Fuks, A., Guild, B. C., Hyafil, F., Kaufman, J. F., Korman, A. J., Kostyk, T. G., Krangel, M. S., Lancet, D., Lopez de Castro, J. A., Mann, D. L., Orr, H. T., Parham, P., Parker, K. C., Ploegh, H. L., Pober, J. S., Robb, R. J. & Shackelford, D. A. (1979) in *The Role of the Major Histocompatibility Complex in Immunobiology*, eds. Benacerraf, B. & Dorf, M. E. (Garland, New York), in press.
3. Cresswell, P., Turner, M. J. & Strominger, J. L. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 1603–1607.
4. Grey, H. M., Kubo, R. T., Cohn, S. M., Poulik, M. D., Cresswell, P., Springer, T. A., Turner, M. & Strominger, J. L. (1973) *J. Exp. Med.* **131**, 1608–1612.

5. Nakamuro, K., Tanigaki, M. & Pressman, D. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 2863–2865.
6. Peterson, P. A., Rask, L. & Lindblom, J. B. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 35–39.
7. Silver, J. & Hood, L. (1974) *Nature (London)* **249,** 764–765.
8. Nathenson, S. G. & Muramatsu, T. (1971) in *Glycoproteins of Blood Cells and Plasma,* eds. Jamiesen, G. A. & Greenwalt, T. J. (Lippincott Co., Philadelphia), pp. 245–262.
9. Parham, P., Alpert, B. N., Orr, H. T. & Strominger, J. L. (1977) *J. Biol. Chem.* **252,** 7555–7567.
10. Shimada, A. & Nathenson, S. G. (1969) *Biochemistry* **10,** 4048–4062.
11. Turner, M. J., Cresswell, P., Parham, P., Strominger, J. L., Mann, D. L. & Sanderson, A. R. (1975) *J. Biol. Chem.* **250,** 4512–4519.
12. Coligan, J. E., Kindt, T. J., Ewenstein, B. M., Uehara, H., Martinko, J. M. & Nathenson, S. G. (1979) *Mol. Immunol.* **16,** 3–8.
13. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A. & Strominger, J. L. (1979) *Nature (London),* in press.
14. Patthy, L. & Smith, E. L. (1975) *J. Biol. Chem.* **250,** 557–564.
15. Udenfriend, S., Stein, S., Bohlen, P., Dairman, W., Leimgruber, W. & Weigele, M. (1972) *Science* **178,** 871–872.
16. Brauer, A. W., Margolis, M. N. & Haber, E. (1975) *Biochemistry* **14,** 3029–3035.
17. Tarr, G. E., Beecher, J. F., Bell, M. & McKean, D. J. (1978) *Anal. Biochem.* **84,** 622–627.
18. Klapper, D. G., Wilde, C. E., III & Capra, J. D. (1978) *Anal. Biochem.* **85,** 126–131.
19. Sauer, R. T., Niall, H. D., Hogan, M. L., Keutmann, H. T., O'Riordan, J. L. H. & Potts, J. T., Jr. (1974) *Biochemistry* **13,** 1994–1999.
20. Gray, W. R. (1967) *Methods Enzymol.* **11,** 333–334.
21. Kulbe, K. D. (1974) *Anal. Biochem.* **59,** 564–573.
22. Summers, M. R., Smythers, G. W. & Orosylan, S. (1973) *Anal. Biochem.* **93,** 624–628.
23. Pisano, J. J. & Bronzert, T. J. (1969) *J. Biol. Chem.* **244,** 5597–5607.
24. Niall, H. D. (1973) *Methods Enzymol.* **27,** 942–1010.
25. Smithies, O., Gibson, D. M., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G. & Ballantyne, D. L. (1971) *Biochemistry* **10,** 4912–4921.
26. Terhorst, C., Robb, R., Jones, C. & Strominger, J. L. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4002–4006.
27. Terhorst, C., Parham, P., Mann, D. L. & Strominger, J. L. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 910–914.
28. Kinquist, K. K. & Lennarz, W. J. (1978) *J. Supramol. Struct.* **8,** 51–65.
29. Joysey, V. C. & Wolf, E. (1978) *Br. Med. Bull.* **34,** 217–222.
30. Parham, P. & Bodmer, W. F. (1978) *Nature (London)* **276,** 397–399.
31. Brodsky, F. M., Parham, P., Barnstable, C. J., Crumpton, M. J. & Bodmer, W. F. (1979) *Immunol. Rev.,* in press.