

Molecular cloning of a human histocompatibility antigen cDNA fragment

(cDNA cloning/positive hybridization-translation/membrane proteins)

HIDDE L. PLOEGH, HARRY T. ORR, AND JACK L. STROMINGER

The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge Massachusetts 02138

Contributed by Jack Leonard Strominger, August 4, 1980

ABSTRACT A clone (pHLA-1) containing HLA-specific cDNA was constructed by reverse transcription of partially purified HLA mRNA from the human lymphoblastoid cell line LKT. The identity of pHLA-1 was established by its ability to hybridize to HLA heavy chain mRNA and by nucleotide sequence analysis. The pHLA-1 cDNA insert (≈ 525 base pairs) corresponds to the COOH-terminal 46 amino acids of an HLA-A, -B, or -C antigen (15 residues from the hydrophobic region and the remainder from the COOH-terminal hydrophilic region), together with a portion of the 3' untranslated region of the mRNA.

The major histocompatibility complex (MHC) is located on the short arm of chromosome 6 in man (1), and includes the HLA system. This genetic region encodes a series of highly polymorphic membrane glycoproteins, the HLA antigens, in addition to components of the complement system, all of which are of crucial importance to a proper functioning of the immune system (2). In addition to the high degree of polymorphism that these MHC products display, the association of certain alleles at HLA loci with a large number of human diseases is of considerable interest (3). The definition of the genetic and mechanistic bases for these disease associations and the evolutionary origin of the polymorphism that may relate to the function(s) of these molecules are most intriguing biological questions.

The classic histocompatibility antigens—i.e., those antigens that are targets for graft rejection—are encoded by the HLA-A, -B, and -C loci. At present, 20, 40, and 8 alleles have been defined at these loci, respectively (4). The number of alleles defined serologically must necessarily represent a minimum estimate of the actual number of alleles present in the human population. A major goal in the structural studies on HLA antigens* has been to relate their complex genetic and serologic properties with structure and function. Amino acid sequence studies of HLA-B7 have provided a molecular description of this antigen (5-9), and comparison of its sequence with partial sequences of other specificities has yielded some insight into the question of regions of allelic variation (10). However, a correlation of primary structure with serologic and functional properties will require the collection of an extensive amount of sequence information. Approaching this problem through determination of the amino acid sequence of a large number of allelic HLA antigens would represent a prodigious effort. Moreover, a purely protein chemical approach cannot address directly some of the fundamental questions concerning the genomic organization of the MHC that are important in understanding the HLA system in detail.

Reported here is the molecular cloning of an HLA-specific cDNA fragment. This cloned fragment provides a tool for the selection and purification of HLA-specific mRNA and can be used as a probe in the isolation of other HLA cDNA clones, thus making the HLA system amenable to analysis by molecular biological methods. Compared to amino acid sequence determination, the determination of nucleotide sequences is a much more rapid and less costly means of accumulating primary structural information. In addition, it is anticipated that this probe will be useful in analyzing the HLA system at the genomic level.

MATERIALS AND METHODS

Growth of Cells and RNA Isolation. The human lymphoblastoid cell line LKT was obtained from the Genetics Laboratory of Oxford University (Oxford, England). The LKT cell line, HLA-A1, -B8, homozygous, was maintained in RPMI-1640 medium (GIBCO) supplemented with 10% fetal calf serum. Cells in middle to late logarithmic phase were used for RNA isolations as described (11).

Enzymes and Other Reagents. Terminal deoxynucleotidyltransferase (terminal transferase) and restriction endonuclease *Hinf*I were obtained from Bethesda Research Laboratories (Rockville, MD). S1 nuclease and the restriction endonuclease *Pst* I were obtained from Boehringer Mannheim. Radiochemicals were obtained from New England Nuclear. A prototype 3'-labeling kit (New England Nuclear, NEK 009) used for terminal transferase-catalyzed labeling of restriction fragments with 3'-deoxy[α - 32 P]ATP (cordycepin) was made available to us through the combined efforts of D. Schwartz and R. di Paolo.

RNA Separation, Cell-Free Translation, and Immunoprecipitation. Cell-free translation in rabbit reticulocyte lysates and analysis of translation products were carried out as described (11). RNA was fractionated on oligo(dT)-cellulose (type 3; Collaborative Research, Waltham, MA) and subjected to size fractionation by electrophoresis on 1.5% agarose gels as described or, alternatively, on a 3% acrylamide gel (acrylamide/bisacrylamide = 20:1) in a device equipped for continuous elution (Heuros BV, Groningen, The Netherlands) (12) using the same buffer system as the agarose gels. Fractions containing HLA heavy chain mRNA activity were localized by translating *in vitro* the size-separated RNA fractions followed by immunoprecipitation. Immunoprecipitation was carried out as described (11), utilizing anti-H—an antiserum that recognizes HLA heavy chains only when not associated with β_2 -micro-

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.

Abbreviations: AMV, avian myeloblastosis virus; bp, basepair(s); ds, double-stranded; MHC, major histocompatibility complex.

* Hereafter, HLA-A, -B, and -C antigens will be referred to as HLA antigens.

globulin (11, 13)—and heat-killed, formalin-fixed *Staphylococcus aureus* Cowan I cells.

cDNA Synthesis. RNA fractions containing HLA heavy chain mRNA activity were used for reverse transcription by using avian myeloblastosis virus reverse transcriptase (a gift of J. Beard). The cDNA was rendered double-stranded by using DNA polymerase I (a generous gift of W. McClure). Reaction conditions for first- and second-strand synthesis were as described by Wickens *et al.* (14). The cDNA was treated with S1 nuclease (15) and, after phenol extraction and gel chromatography on Sephadex G-100, the nucleic acids eluting in the void volume were ethanol precipitated. Homopolymer dC tails were added by using terminal transferase at 37°C for 15 min (16). The tailed cDNA was then subjected to size separation on a 6% polyacrylamide gel, and the fraction with an apparent size >600 base pairs (bp) was recovered (17). This fraction was annealed with pBR322 that had been linearized with *Pst* I and tailed with dG (16). Chimeric plasmids were used to transform *Escherichia coli* χ 1776 (18). Colonies that were resistant to tetracycline at 12 μ g/ml were selected and screened for the presence of HLA sequences by using the mRNA selection assay described below.

These experiments were carried out in compliance with the National Institutes of Health guidelines on recombinant DNA research.

RNA Selection on Nitrocellulose Filters. The method of Ricciardi *et al.* (19) was used essentially as described but with the following modifications, based in part on the "dot-blot" procedure of Kafatos *et al.* (20). Recombinant DNAs to be tested were obtained from 10-ml cultures of *E. coli* χ 1776 harboring the relevant plasmids, amplified by addition of chloramphenicol (50 μ g/ml) at OD₅₅₀ = 1.0. From these cultures, cleared lysates were prepared by sucrose/lysozyme/EDTA treatment followed by lysis with NaDodSO₄ and precipitation of chromosomal materials with KCl. After two ethanol precipitations, the resulting pellet was taken up in 100 μ l of 10 mM Tris-HCl, pH 7.5/2 mM EDTA, and 1 M NaOH was added to a final concentration of 0.33 M. After 10 min at room temperature, the solution was neutralized with 1 M HCl, diluted 1:2 with 2 M ammonium acetate, and spotted onto a square (approximately 10 mm²) of nitrocellulose. This was conveniently done by supporting a 28-mm diameter nitrocellulose filter (Schleicher & Schuell, BA 85), marked into 10-mm² squares with a soft pencil, on a fritted glass disk and applying a slight vacuum by means of an aspirator. After sample application, the filters were rinsed with 0.75 M NaCl/0.075 M sodium citrate (5 \times standard saline citrate), air dried, and baked under reduced pressure at 65°C for at least 2 hr. The filters were then cut into small squares with a sterile razor blade and prehybridized for 2 hr at 42°C in a solution containing 50% formamide, 10 mM piperazine-*N,N'*-bis(2-ethanesulfonic acid) (pH 6.8), 0.4 M NaCl, 5 μ g of yeast tRNA (Sigma) per ml, and 10 μ g of poly(A) (Collaborative Research) per ml. The prehybridization solution was then replaced by a similar solution containing poly(A)⁺RNA from LKT cells instead of poly(A), at a concentration of 200–800 μ g/ml, with no added yeast tRNA. Hybridization was carried out for 12–16 hr at 42°C. After hybridization, the filters were washed and the bound RNA was recovered as described (19). The eluted RNA was used to program a 25- μ l *in vitro* translation reaction (11). After translation, 5 μ l of the translation products were run on a 7–15% gradient gel directly (21), and the remaining 20 μ l was used for immunoprecipitations.

3'-End Labeling and DNA Sequence Analysis. 3'-End labeling was carried out according to the instructions supplied

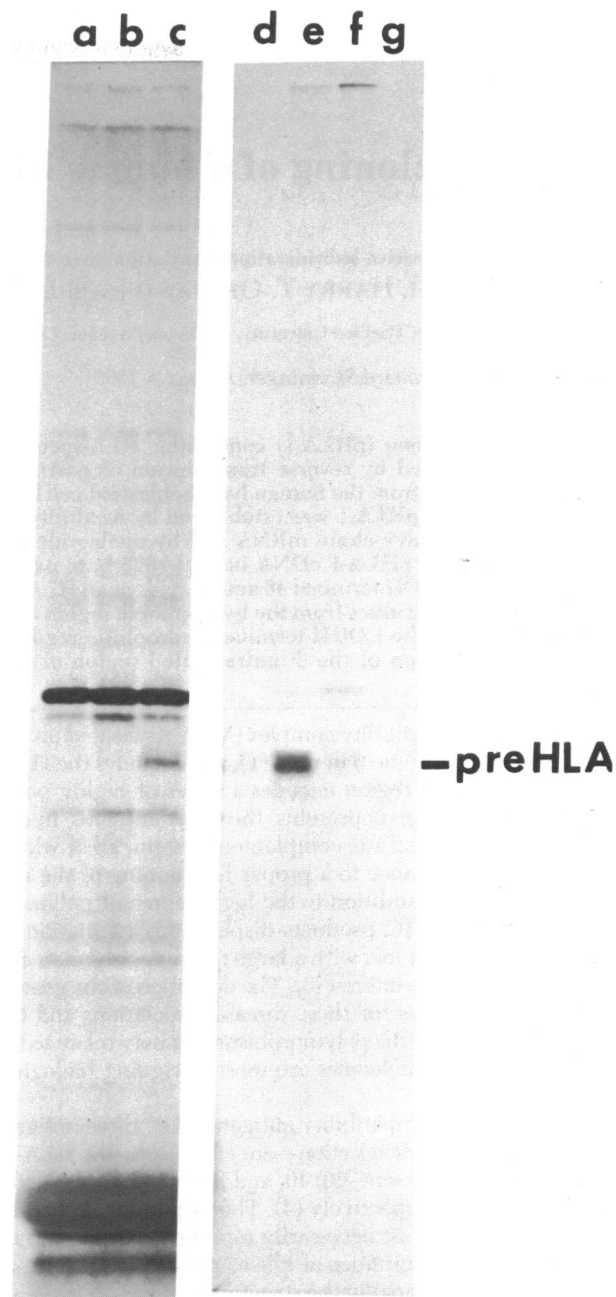


FIG. 1. Immunoprecipitation of translation products programmed by pHLA-1 selected mRNA. Recombinant plasmid DNAs immobilized on nitrocellulose were used to select mRNAs from total LKT poly(A)⁺RNA. Translation products were analyzed by immunoprecipitation and NaDodSO₄/polyacrylamide gel electrophoresis. (Left) Total translation products of RNA selected by two recombinant plasmids not containing an HLA-specific insert (lanes a and b) and pHLA-1 (lane c). (Autoradiographic exposure time, 12 hr.) (Right) Anti-H immunoprecipitations of translation products: lane d, products from lane b; lane e, products from lane c (pHLA-1); lane f, products from total RNA (a faint band at the position of preHLA did not reproduce clearly here); lane g, products using no RNA. (Autoradiographic exposure time 4 hr.) Plasmids used for the selections in lanes a and b have inserts that have not been characterized. The heavy band of radioactivity in lanes a, b, and c migrating above pre-HLA is an artifact of translation that is also found when no RNA is added (11).

by the manufacturer (New England Nuclear) utilizing [α -³²P]cordycepin and terminal transferase. Labeled DNA fragments were purified by gel electrophoresis; sequence deter-

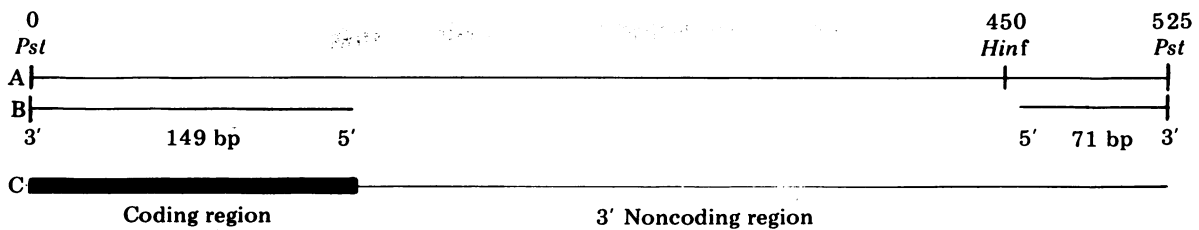


FIG. 2. Diagram of the pHLA-1 525-bp insert. (A) Location of the *HinfI* site relative to the *Pst* I sites. (B) Portions of the two pHLA-1 *HinfI* fragments analyzed (3'-to-5'). (C) 5'-to-3' depiction of the coding and 3' noncoding regions of HLA mRNA complementary to the pHLA-1 insert.

mination was according to the method of Maxam and Gilbert (17).

RESULTS

Selection of pHLA-1, an HLA-Specific cDNA-Containing Plasmid. Poly(A)⁺RNA from the LKT cell line (HLA-A1, -B8 homozygous) was enriched for HLA mRNA activity by electrophoretic fractionation (11). This resulted in a 20- to 40-fold purification as assayed by cell-free translation. After reverse transcription of HLA mRNA-containing fractions, size-selected double-stranded (ds) cDNA was obtained. This ds cDNA was then inserted into the *Pst* I site of pBR322 by using the G-C tailing method (16). *E. coli* χ 1776 was transformed with the recombinant plasmids and approximately 200 tetracycline-resistant colonies were obtained. The resulting transformants were screened individually for the presence of HLA sequences by the ability of their DNA immobilized on nitrocellulose filters to retain HLA mRNA.

Extensive purification of plasmid DNA was unnecessary because immobilized DNA from ethanol-precipitated cleared lysates of the transformants was able to hybridize specifically to mRNAs (Fig. 1). In this fashion, relatively large numbers of transformants were screened without the need for tedious and time consuming CsCl banding of recombinant plasmids. The RNA-dependent background observed in the translation assays (apparently due to nonspecific binding of some mRNAs) did not prevent an unambiguous identification of unique bands present above background. The pattern of background bands for any given preparation of RNA in fact was quite constant.

Those recombinant plasmids that were able to retain mRNAs that translated into protein in the size range of the HLA heavy chain precursor were subjected to further analysis. The trans-

lation products obtained using these recombinants were immunoprecipitated with anti-H. This antibody is specific for HLA heavy chains (11, 13) and has no known cross reactions with other proteins; the immunogen was an extensively purified HLA heavy chain preparation. One clone (pHLA-1) was obtained that retained with very high efficiency an mRNA that was translated into a polypeptide of pre-HLA size (Fig. 1, lane c). This band was immunoprecipitable with anti-H serum (Fig. 1, lane e) and, therefore, this clone was considered to be an HLA-specific cDNA clone.

Sequence Analysis of the pHLA-1 cDNA Insert. Direct sequence analysis of the insert confirmed the presence of HLA-specific cDNA. The insert of pHLA-1 was excised with *Pst* I and found to be approximately 525 bp long. For sequence analysis, pHLA-1 was cut with *Pst* I and the 3' termini were labeled with [α -³²P]cordycepin in a terminal transferase-catalyzed reaction (refs. 22 and 23; A. M. Maxam and D. Schwartz, personal communication). The end-labeled DNA was then cleaved with *HinfI*. Because the pHLA-1 insert contains a single *HinfI* site (Fig. 2A), this cleavage results in the generation of two uniquely end-labeled fragments from the insert which were easily separated on and then recovered from a 6% polyacrylamide gel for sequence analysis. In this way, sequence information was obtained from both ends of the pHLA-1 insert. Analysis of the nucleotide sequences obtained indicated that the insert did indeed contain HLA sequence information (Fig. 3).

The partial nucleotide sequence of the strand complementary to the sequence from the large *HinfI* fragment (Fig. 2) yielded, in one of its three possible reading frames, an amino acid sequence identical to the COOH-terminal portion of HLA-B7 (5-7). Distal to the G-C tail (15 bp) 119 nucleotides were

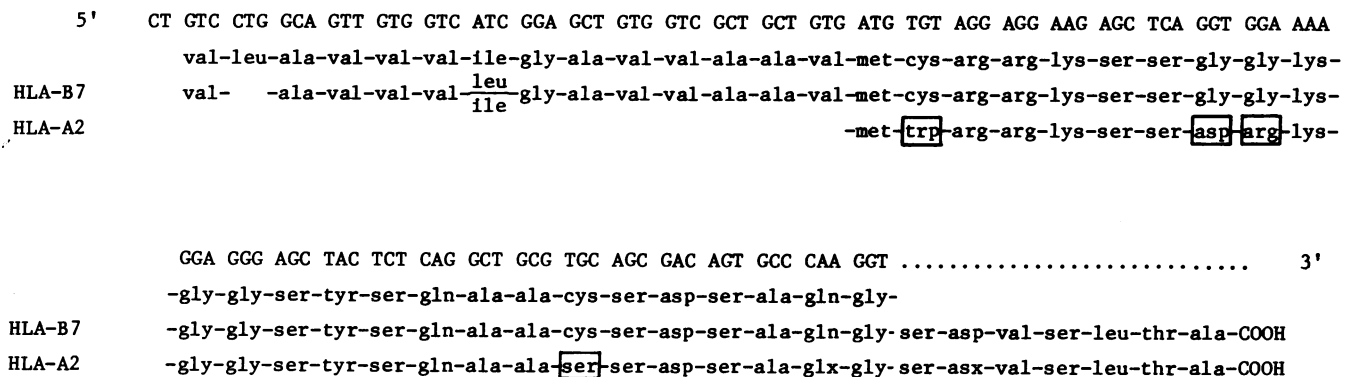


FIG. 3. Partial nucleotide sequence of pHLA-1. A *Pst* I digest of pHLA-1 was labeled with [α -³²P]cordycepin and terminal transferase. The labeled products were digested with *HinfI* and separated on a 6% polyacrylamide gel. The labeled *HinfI* fragments unique to pHLA-1 were recovered from the gel and subjected to sequence analysis. The partial nucleotide sequence of the strand complementary to that sequenced from the large *HinfI* fragment and the amino acid sequence deduced from it are shown. In addition, the protein sequences from the COOH termini of HLA-B7 and HLA-A2 are shown (5-7). Differences between the cDNA-deduced sequence and the protein sequences are indicated by the boxed residues.

identified. This sequence begins within the coding region of the intramembranous hydrophobic portion of HLA (6, 7, 24, 25) and extends into the COOH-terminal hydrophilic region for 24 amino acid residues.

The remaining unsequenced portion of the insert, therefore, must consist of the coding region for the remaining seven COOH-terminal amino acids of HLA and a portion of the 3' untranslated region of the mRNA (Fig. 2C). Nucleotide sequence data from the small (76 bp) *Hinf*I fragment (Fig. 2) failed to reveal either the poly(A) tail or the hexanucleotide (A-A-T-A-A-A) generally found 15–30 bp upstream from the poly(A) tail (26). Thus, it is likely that the pHLA-1 cDNA insert does not extend to the 3' end of the HLA mRNA. Based on an HLA chain of 343 residues (5–7, 9, 11) including the NH₂-terminal signal sequence, the mRNA coding region would be 1029 nucleotides long. Size estimates of HLA mRNA indicate it to be approximately 1650 bases (unpublished data). Thus, the total noncoding portion (5' and 3') of HLA mRNA could be as much as 620 nucleotides.

DISCUSSION

Starting with partially purified mRNA, an HLA cDNA clone (pHLA-1) has been isolated by established techniques. pHLA-1 was shown to contain an HLA-specific cDNA insert by its ability to hybridize specifically to HLA mRNA as judged by cell-free translation and immunoprecipitation. Because only limited NH₂-terminal amino acid sequence information is available for the LKT HLA antigens (HLA-A1 and HLA-B8) (27, 28), the extensive amino acid sequences available for the HLA-A2 and HLA-B7 antigens were used for comparison with the amino acid sequence deduced from pHLA-1. At the 38 amino acid positions available for comparison, complete identity with HLA-B7 was found (Fig. 3). This unambiguously identifies pHLA-1 as an HLA-specific cDNA clone and suggests that the COOH-terminal region including the intramembranous hydrophobic stretch of HLA antigens is highly conserved between HLA specificities. Because the amino acid sequence of the HLA-A2 COOH-terminal hydrophilic portion differs in four positions from that of HLA-B7 (and thus from the sequence deduced from pHLA-1) (5) (Fig. 3), it seems warranted to speculate that the cDNA clone pHLA-1 corresponds to the COOH-terminal portion of the LKT HLA-B locus antigen, HLA-B8, rather than to the LKT HLA-A locus antigen, HLA-A1.

Two aspects of the methods used in constructing and sequencing the cDNA clone deserve further comment. First, it is clear that for mRNA selection the recombinant DNA molecules need not be very pure and that small cultures (10 ml) are sufficient to generate detectable signals in the RNA selection assay. No difficulties were observed in immobilizing this relatively impure DNA to nitrocellulose filters or in translation of bound and eluted mRNAs. It is likely that most procedures for plasmid "mini-preps" (e.g., ref. 29) will yield DNA of sufficient purity for use in this assay, thus enabling screening of large numbers of recombinants in reasonable periods of time.

In the determination of the partial sequence of pHLA-1, the use of [α -³²P]cordycepin and terminal transferase proved to be particularly useful for labeling the 3' ends of the *Pst* I excised fragment, allowing determination of the nucleotide sequence at both ends of the insert. Because this method is well suited for labeling restriction fragments with protruding 3' ends, it should prove to be of general use for the rapid sequence determination of segments of cDNA cloned into the *Pst* I site of pBR322, eliminating the need for extensive restriction mapping prior to sequencing.

Now that an MHC-specific probe is available, some of the

questions posed by serological and genetic analyses of the MHC can be approached in molecular terms. In particular, the prospect of generating complete amino acid sequences for a number of allelic HLA specificities should allow a more precise definition of regions that are involved in allelic variability. Second, from such sequence analyses, some of the uncertainties still present in the amino acid sequence can be resolved, such as the overlap between the papain-solubilized NH₂-terminal portion and the hydrophobic segment, and the firm establishment of the sequence of the intramembranous hydrophobic region. More important, the genomic structure of HLA genes can now be probed directly, which should provide valuable information on the genetic organization of this region. Given the high degree of sequence homology that exists between MHC products from homologous loci of the different animal species examined to date (30), MHC probes such as the one described in this paper might be of considerable use in the comparative analysis of MHC structure.

Note Added in Proof. The presence of an *Sst* I site (G-A-G-C-T⁺C) 63 base pairs from the 5' end of the pHLA-1 coding region has been utilized to obtain additional nucleotide sequence information. The sequence TCT GAT GTG TCT CTC ACA GCT TGA is contiguous with the 3' end of the pHLA-1 coding sequence shown in Fig. 3. The amino acid sequence deduced from this additional stretch completes the coding region of pHLA-1 and matches exactly the amino acid sequence of HLA-B7 up to and including the COOH terminus (Fig. 3). The COOH-terminal alanine codon (GCT) is followed by the stop codon TGA, indicating the lack of COOH-terminal proteolytic processing *in vivo*.

We acknowledge the suggestions and advice from the members of the Gilbert, Kafatos, Losick, Pero, and Ptashne laboratories. In particular we thank D. Alexandraki, A. Bothwell, G. Lee, C. Moran, T. Roberts, D. Schwartz, G. Thireos, R. Tizard, and K. Zinn. We thank Willem Roskam for making available his preparative gel electrophoresis apparatus and instructing us in its use. This work was supported by National Institutes of Health Research Grants AI-10736 and AM-13230. H.T.O. is a Special Fellow of the Leukemia Society of America.

1. Francke, U. & Pellegrino, M. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1147–1151.
2. Bach, F. H. & Van Rood, J. J. (1976) *N. Engl. J. Med.* **295**, 806–813.
3. Dausset, J. & Svejgaard, A., eds. (1977) *HLA and Disease* (Munksgaard, Copenhagen).
4. Dick, H. (1980) *Immunol. Today*, **1**, 1–7.
5. Robb, R. J., Terhorst, C. & Strominger, J. L. (1978) *J. Biol. Chem.* **253**, 5319–5324.
6. Ploegh, H. L., Orr, H. T., Robb, R. J. & Strominger, J. L. (1978) in *Biological Markers of Neoplasia: Basic and Applied Aspects*, ed. Riddon, H. (Elsevier/North-Holland, New York), pp. 201–211.
7. Robb, R. J. (1977) Dissertation (Harvard Univ., Cambridge, MA).
8. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A. & Strominger, J. L. (1979) *Nature (London)* **282**, 266–270.
9. Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. L. (1979) *Biochemistry* **18**, 5711–5720.
10. Orr, H. T., Lopez de Castro, J. A., Parham, P., Ploegh, H. L. & Strominger, J. L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4395–4399.
11. Ploegh, H. L., Cannon, L. E. & Strominger, J. L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2273–2277.
12. Wieringa, B., Mulder, J., Van der Ende, A., Bruggeman, A., Ab, G. & Gruber, M. (1978) *Eur. J. Biochem.* **89**, 67–69.
13. Krangel, M. S., Orr, H. T. & Strominger, J. L. (1979) *Cell* **18**, 979–991.

14. Wickens, M. P., Buell, G. N. & Schimke, R. T. (1978) *J. Biol. Chem.* **253**, 2483–2495.
15. Roberts, T. M. & Lauer, G. D. (1979) *Methods Enzymol.* **68**, 473–492.
16. Nelson, T. & Brutlag, D. (1979) *Methods Enzymol.* **68**, 41–50.
17. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
18. Katcoff, D., Nudel, U., Zevin-Sonkin, D., Carmon, Y., Shani, M., Lehrach, H., Frischauf, A. M. & Yaffe, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 960–964.
19. Ricciardi, R. P., Miller, J. S. & Roberts, B. E. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4927–4931.
20. Kafatos, F. C., Jones, C. W. & Efstratiadis, A. (1979) *Nucleic Acids Res.* **7**, 1541–1552.
21. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
22. Olson, K. & Harvey, C. (1975) *Nucleic Acids Res.* **2**, 319–325.
23. Tu, P.-P. D. & Cohen, S. N. (1980) *Gene* **10**, 177–183.
24. Springer, T. A. & Strominger, J. L. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2481–2485.
25. Goldman, D. W., Pober, J. S., White, J. & Bayley, H. (1979) *Nature (London)* **280**, 841–843.
26. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
27. Bridgen, J., Snary, D., Crumpton, M. J., Barnstable, C., Goodfellow, P. & Bodmer, W. F. (1976) *Nature (London)* **261**, 200–205.
28. Ballou, B., McKean, D. J., Freedlender, E. F. & Smithies, O. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 4487–4491.
29. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
30. Klein, J. (1979) *Science* **203**, 516–521.