Molecular Medicine 4: 443-453, 1998

Molecular Medicine

Original Articles

Genotyping HIV-1 and HCV Strains by a Combinatorial DNA Melting Assay (COMA)

Leondios G. Kostrikis, Sunny Shin, and David D. Ho

Aaron Diamond AIDS Research Center, The Rockefeller University, New York, New York, U.S.A.

Communicated by D. Ho. Accepted May 21, 1998.

Abstract

Background: Human immunodeficiency virus type 1 (HIV-1) and hepatitis C virus (HCV) strains can be genetically classified into genetic lineages known as genetic types or subtypes according to phylogenetic analyses of complete or partial nucleotide sequences of their genomes. The genetic classification of HIV-1 and HCV strains has important implications for the development of globally effective vaccines and for the management of patients.

Materials and Methods: A new method, termed combinatorial DNA melting assay (COMA), allows rapid accessing of comparative genetic information between related DNA sequences, making it possible to rapidly and accurately genotype unknown HIV-1 and HCV strains. COMA is mainly based on the differential melting properties of long DNA heteroduplexes. Combinatorial arrays of DNA heteroduplexes are formed when captured PCR-amplified reference DNA with known nucleotide sequences are combined with solution-phase complementary and antigenically labeled DNA with unknown sequences. Genetic divergence between the known and the unknown sequences is inferred as the experimentally derived melting curves of the two strands of the DNA heteroduplexes increasingly diverge.

Results: COMA was successfully applied to the genetic classification of HIV-1 and HCV strains into phylogenetic lineages or subtypes.

Conclusions: Use of this assay should accelerate current efforts to understand the global molecular epidemiology of HIV-1 and HCV and may extend to the genetic characterization of other genetically diverse infectious pathogens associated with numerous diseases.

Introduction

Simple and accurate methods of genotyping human pathogens are increasingly in demand. Infectious agents with extensive sequence heterogeneity, such as human immunodeficiency virus type 1 (HIV-1) or hepatitis C virus (HCV), often require genetic characterization, typically carried out by conventional nucleic acid sequencing.

Address correspondence and reprint requests to: Dr. David D. Ho, Aaron Diamond AIDS Research Center, The Rockefeller University, 455 First Avenue, New York, NY 10016, U.S.A. Phone: (212) 725-0018; Fax: (212) 725-1126; E-mail: dho@adarc.org

Classification of HIV-1 into distinct phylogenetic lineages, known as sequence subtypes, or "clades," is based on phylogenetic analysis of *gag* and *env* gene sequences (1). Classification of HCV strains into major types and subtypes is based on sequence analysis of the 5′ noncoding region (NC), the core (C), and the envelope 1 (E1) and nonstructural protein 5 (NS5) coding regions (2–7). The large database of HIV-1 and HCV sequences now available presents an opportunity to develop new methods for rapidly determining the genetic relationship between unknown viral isolates and known reference sequences. In one

approach, an unknown HIV-1 *env* sequence can be accurately subtyped according to the electrophoretic mobility of a set of DNA heteroduplexes formed with individual reference sequences (8). Although this type of heteroduplex mobility assay is now in wide use, its requirement of analytical gel electrophoresis makes it considerably labor intensive. We describe here a new method, termed combinatorial melting assay (COMA), that measures the differential melting properties of DNA heteroduplexes with varying degrees of sequence divergence between the two strands. COMA produces a final colorimetric immunoassay-based readout that allows the simultaneous comparison of an unknown sequence with multiple reference sequences. We designed a COMA that could determine the genetic subtypes of HIV-1 or HCV by efficiently and accurately analyzing sequences encoding the *env* or the *C* and *E1* (*C-E1*) (5) coding regions, respectively.

Materials and Methods

HIV-1 and HCV Strains

HIV-1 samples were provided by E. L. Delwart (8), M. L. Kalish, and B. Hahn (9) as blinded unknowns. The first two or three letters of the unknown sample name refer to the country of origin: BZ or BR, Brazil; DJ, Djibouti; EG, Egypt; HT, Haiti; KE, Kenya; MW, Malawi; POC, United States Navy servicemen infected outside the United States; RO, Romania; RU, Russia; RW, Rwanda; SM, Somalia; SN, Senegal; TH, Thailand; TZ or TAN, Tanzania; UG, Uganda; US, United States. Samples consisted of extracted genomic DNA from patients' peripheral blood mononuclear cells (PBMC) (E. L. D), extracted DNA from primary PBMC cultures (B. H.), or primary polymerase chain reaction (PCR)-amplified products derived from patients' PBMC (M. L. K).

HCV plasma samples were provided by M. Busch (Irwin Memorial Blood Centers, San Francisco, CA), J. Kolberg, M. Urdea, and A. Weiner (Chiron Corp., Emeryville, CA). The first two letters of the reference sample name refer to the country of origin: US, United States; IT, Italy; FR, France. The next three letters refer to the place of origin: CHI, Chiron Corp.; IRW, Irwin Memorial Blood Centers.

The AIDS Research and Reference Reagent Program (AIDS Program, NIAID, NIH) provided the subtype reference sequences and the plasmids containing RW20, IC144, SF170, BR20, TH14, MA959, IN868, BR25, UG21, UG38, UG46, TH22, TH06, BZ126, BZ162, BZ163, RU131, LBV21-7, VI525, CA13 and VI557 HIV-1 *env* genes. CYHO401, CYHO111, CYHO503, CYHO481, and CYHO271 were obtained from the Aaron Diamond AIDS Research Center (ADARC) repository (10). The remaining subtype references used were PCR-amplified sequences obtained from samples whose genetic subtype had been previously determined by COMA.

HIV-1 and HCV DNA Heteroduplex Melting Curves

For HIV-1, reference "capture" single-stranded DNA molecules encoding the gp120 C2 to V3 region were amplified from known sequences (either genomic DNA or plasmids) usually by two rounds (one symmetric and one asymmetric) of PCR (PCR primers: LK1, 5′-CCAATTCCCATACATTATTGTGCCCCGG CTGG and LK2, 5′-TTACAGTACAAAAATTCCCCTC-CACAATTAAA). The PCR primers were designed to amplify sequences (~530 bases in size, depending on the length polymorphism of each strain) from all the genetic subtypes of HIV-1 within the M group (A through H). The first round of amplification used 1 μg of genomic DNA, 20 pmoles of each PCR primer, and 5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus) in a 100-μl solution at a final concentration of 250 nM dNTP and 1.5 mM $MgCl_2$. The thermocycling conditions were 1 cycle at 94°C for 2 min, 35 cycles at 94°C for 20 sec, 55°C for 30 sec, and 72°C for 30 sec, and 1 cycle at 72°C for 1 min. The second round of PCR used approximately 0.5–1.0 pmoles of template from the first amplification, 1–10 pmoles of LK2, and 70–100 pmoles of LK1B (5′ biotinylated LK1 primer) in conditions identical to the first round. Alternatively, the second round of PCR may utilize the same amount of LK2 and LK1B primers (symmetric PCR).

For HCV, reference capture single-stranded DNA molecules encoding a C-E1 region (491 nucleotides long) were amplified from known HCV sequences cloned in plasmids by an asymmetric PCR (PCR primers: LK3, 5′-GGGGAC-CAGTTCATCATCATATCCCATGCC-3′ and LK4, 5′-GCAACAGGGAATTTGCCCGGTTGCTCTT-TCTC-3′) using 70–100 pmoles of 5′ biotinylated LK4 and 1–10 pmoles of LK3. All other PCR reagents and thermal cycling conditions were the same as for HIV-1. The unknown HIV-1 single-stranded DNA "detection" molecules were gen-

erated from genomic DNA isolated from HIV-1-infected peripheral blood mononuclear cells by two rounds of nested PCR. Primers used in the first PCR round were 20 pmoles each of ED5 and ED12 (8). The second PCR round used inner primers LK1 (1–10 pmoles) and LK2 (70–100 pmoles), 2 mM each of dGTP, dCTP, and dATP, 1.9 mM dTTP, and 0.1 mM digoxigenin-11-dUTP (Boehringer Mannheim). All other PCR reagents and thermal cycling conditions were the same as previously described. The HCV single-stranded DNA "detection" molecules were generated by reverse transcription of plasma HCV RNA and two rounds of PCR. HCV RNA was extracted from plasma (100–200 μl) with the QIAamp viral RNA purification protocol (QIagen), and extracted RNA was diluted in 50 μl RNase-free H_2O. RNA (3 μl) was reverse transcribed into cDNA and amplified by PCR (RT-PCR). One hundred pmoles of primer LK5 (5′-CCTTCGCCCAGTTCCCCACCATGGAGAAATACGC) was used in each reverse transcription reaction, and 100 pmoles of primer LK6 (5′-GCCGACCTCATGGGGTACATCCCGCTCGTCGG) was added for the first-round PCR. Sequences encoding a C-E1 region (491 nucleotides long) were amplified from each RT-PCR by a secondary symmetric or asymmetric PCR. Each 100 μl of secondary PCR contained 5 μl of first-round RT-PCR products, LK3 and LK4 primers (70–100 pmoles each), 2 mM each of dGTP, dCTP, and dATP, 1.9 mM dTTP, and 0.1 mM digoxigenin-11-dUTP in 100 μl reaction volume. All other PCR reagents and thermal cycling conditions were the same as for symmetrical PCR. The unknown HCV DNA molecules were then placed across the known HCV DNA in rows of the assay plate, thus allowing the formation of heteroduplexes between each unknown DNA and each reference pool.

Combinatorial Melting Assay

For the colorimetric assay, 96-well streptavidin-coated microplates were used (Pierce). The capture ssDNA was denatured for 5 min at 95°C, centrifuged to pull down any condensate, and put into ice. Five pmoles of the denatured capture ssDNA was put into each microwell. The solution in the well was brought up to 30 μl to ensure full coverage of the well bottom. The plate was incubated for 5 min on ice to ensure that the biotin-labeled probes bound to the streptavidin-coated well and the duplex DNA remained denatured after the 5-min heating at 95°C. If the reference DNA was derived from symmetric PCR, biotinylated ssDNA was captured into each streptavidin-coated microwell for 5 min at room temperature, denatured for 5 min at 95°C, and washed twice with 10 mM Tris-Cl and 0.1 mM EDTA (pH 8.0) to remove the complementary unbiotinylated ssDNA. To block any remaining streptavidin and nonspecific sites, the wells were washed three times with 200 μl TBS blocking buffer (Pierce) containing 2 mM biotin. For the first wash, the buffer was incubated in the well for 10 min. The wells were subsequently equilibrated for hybridization by washing with 100 ml "rapid-hyb" buffer (Amersham). Fifty microliters of hybridization buffer was then pipetted into each well. Usually eight microwells per plate (last eight-well row) were used as negative controls (no "detection" ssDNA was placed into them). For a positive control, detection ssDNA of a known subtype was placed across an eight-well row. The detection ssDNA was then denatured for 5 min at 95°C, and 10 pmoles of this ssDNA was pipetted into each well. The hybridization step was carried out at 75 ± 1°C and 79 ± 1°C for HIV-1 and HCV, respectively, for 2 hr on a thermal block (USA Scientific) or on a customized aluminum holder on a Perkin-Elmer 9600 Thermocycler under constant temperature. After incubation, the plate was drained and stringently washed. For the first stringent washing, 200 μl of 2× SSC, 0.1% (w/v) SDS was pipetted into each well and incubated for 20 min at room temperature. For the next two washings, 200 μl of 1× SSC, 0.1% (w/v) SDS was pipetted into each well and incubated for 15 min at the hybridization temperature. The plate was then put on ice and washed two times with 200 μl blocking buffer (Pierce). For the second wash, the plate was incubated for 30 min to block nonspecific sites. For color development, monoclonal anti-DIG antibody covalently attached to alkaline phosphatase (Boehringer Mannheim) was diluted 1:2500 in 100 mM Tris-Cl and 150 mM NaCl (pH 7.5). Then 200-μl aliquots of the dilution mixture were pipetted into each well and incubated for 30 min at room temperature to ensure that the digoxigoxin-labeled DNA was recognized by the antibody. After incubation, all wells were washed several times before color development using the AMPAK kit (DAKO). Absorbance densities were measured at 490 nm in a 96-well spectrophotometer (Biolumin 960, Molecular Dynamics). For each microwell, the relative absorbance density was derived as follows: [absorbance density] − [average absorbance density from eight "negative controls" included in each

96-well plate]. For each unknown sample, comparative absorbance ratios corresponding to each of the subtype-reference pools were derived by dividing the relative absorbance densities for each microwell by the highest value. The subtype-reference pool that gives the highest relative absorbance (i.e., comparative absorbance ratio = 1.00) indicates the correct genetic subtype for the unknown sample provided that the other comparative absorbance ratios corresponding to the remaining subtype-reference pools are ≤0.50.

HCV Phylogenetic Analysis

To verify the results obtained from COMA, the DNA region corresponding to the HCV RNA-binding nucleocapsid and E1 envelope glycoprotein was amplified by RT-PCR, cloned into pCR II bacterial plasmid, and sequenced with an Applied Biosystems automated DNA sequencer. The genetic subtypes were determined by phylogenetic analyses using CLUSTAL W 1.6. GenBank accession numbers for nucleotide sequences are pending.

Results

Reference HIV-1 and HCV Melting Curves

Important to COMA development is the establishment of reference melting curves for DNA heteroduplexes with increasing sequence diversity. Such melting curves were experimentally derived by varying two parameters: the hybridization temperature and the sequence diversity between the two annealing strands. To generate HIV-1 melting curves (Fig. 1A), the *env* genes of previously characterized HIV-1 *env* clones were amplified by polymerase chain reaction to produce a collection of complementary, reference "capture" and "detection" single-strand DNA molecules required for heteroduplex formation (Table 1). The stability of each DNA duplex was determined by absorbance measurements of the digoxigenin-labeled DNA strands using a colorimetric immunoassay. Three such melting curves that represent the approximate intra- and intersubtype diversity boundaries (Fig. 1A) were obtained by multiple absorbance measurements of the homoduplex and heteroduplexes. To generate HCV melting curves, genomic fragments encoding the carboxy terminus of the RNA-binding nucleocapsid protein (C) and the amino terminus of the E1 envelope glycoprotein were reverse transcribed, amplified by PCR, and cloned in bacterial plasmids, and their genetic subtypes were determined by phylogenetic analysis. The melting of HCV heteroduplexes was measured at higher temperatures because of their intrinsically higher G/C composition, resulting in a higher DNA melting temperature (Table 1). Four melting curves that represent the approximate intra- and intersubtype and intergenotype diversity boundaries were obtained by multiple absorbance measurements of the homoduplex and heteroduplexes (Fig. 1B).

The right panel of Figure 1A presents the phylogenetic tree of 18 representative HIV-1 *env* sequences; their genetic subtypes were previously determined according to established criteria (9). In this particular phylogenetic tree, we included members of the major sequence subtypes of HIV-1 group M (subtypes A to G) only. As can be seen, the sequence divergence between any two strains within any of the subtypes (obtained by summing the horizontal lengths connecting the two strains, using the scale at the bottom of the tree) ranged from about 4% to 14%; and the divergence between strains from any two different subtypes ranged from about 15% to 28%. Depending on the total number and the diversities of HIV-1 *env* sequences analyzed by similar phylogenetic analysis, the intra- and intersubtype diversities may vary by about 3–5%. The left panel of Figure 1A presents the results of COMA. The three melting curves obtained for HIV-1 define the boundaries between hybridization signals from intrasubtype and intersubtype heteroduplexes. The upper melting curve presents hybridization signals that resulted from submitting to COMA DNA homoduplexes derived from single HIV-1 strains (0% divergence). The middle melting curve presents hybridization signals that resulted from submitting to the assay DNA heteroduplexes derived from strains from within the same HIV-1 subtype (15% divergence). The lower melting curve presents hybridization signals that resulted from submitting to the assay DNA heteroduplexes derived from strains of different HIV-1 subtypes (30% divergence). The normalized absorbance data used to construct these curves are from absorbance measurements presented in Table 1. The T_m for the HIV-1 DNA homoduplex curve was 75 ± 0.5°C; the T_m for the intrasubtype DNA heteroduplex curve was 74 ± 0.5°C; and the T_m for the intersubtype DNA heteroduplex curve was 72 ± 0.5°C. At 75 ±0.5°C, the absorbance ratio of homoduplex to intersubtype heterodu-

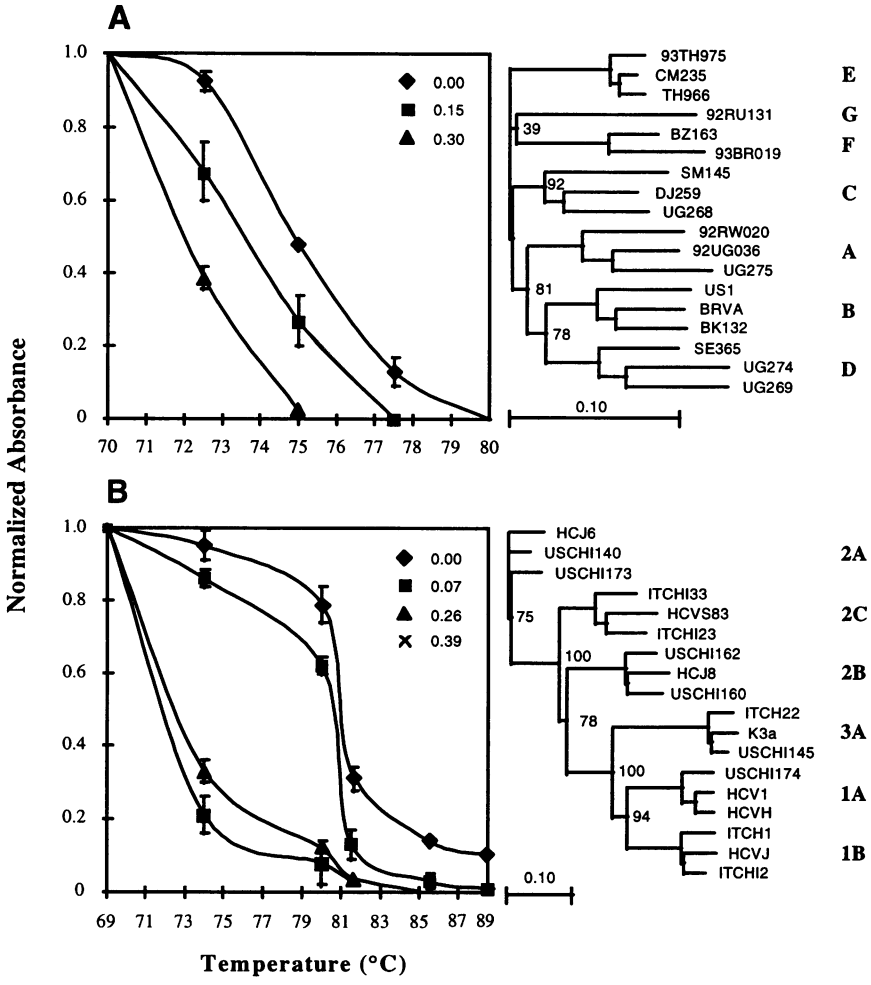

Fig. 1. Representative melting curves of HIV-1 (A) and HCV (B) heteroduplexes used in combinatorial melting assays (left) and representative phylogenetic trees of reference viral sequences (right). Normalized absorbance densities at 490 nm are plotted against temperature measurements for constructed homoduplexes and heteroduplexes of increasing diversity. Normalized absorbance measurements derived for HIV-1 and HCV were normalized by dividing the corresponding relative absorbance measurements (see Table 1) by the highest absorbance value at 70°C for HIV-1 and 69°C for HCV. The error bars represent the standard deviation of multiple capture assays. (A) HIV-1 *env* duplexes: solid diamond = 0% divergence; solid square = 15% divergence; solid triangle = 30% divergence. (B) HCV *C-E1* duplexes: solid diamond = 0% divergence; solid square = 7% divergence; solid triangle = 26% divergence; X = 39% divergence. The numbers next to each of the major phylogenetic "forks" refer to bootstrap values. The error bars represent the standard deviation from independent experiments (n = 5).

plex was greater than 10 to 1, and at higher temperatures the intersubtype heteroduplex signals disappeared entirely. At 77 ± 0.5°C, the absorbance ratio of homoduplex to intrasubtype heteroduplex was greater than 10 to 1, and at higher temperatures the intrasubtype heteroduplex signals disappeared entirely. Therefore, we conclude that thermal melting of HIV-1 heteroduplexes at 75° to 77.5°C should be sufficient to distinguish intrasubtype from intersubtype heteroduplexes.

Similarly, the right panel of Figure 1B presents the phylogenetic tree of 18 representative HCV *C-E1* coding regions, some of which have been previously documented (3–4) and others of which, indicated by an asterisk, were determined as part of this study (see Materials and Methods). As can be seen, the sequence divergence between any two strains within any of the subtypes (obtained by summing the horizontal lengths connecting the two strains, using the scale at the bottom of the tree) ranged from about 5% to

Table 1. Relative optical densities of HIV-1 and HCV homoduplexes and heteroduplexes derived by COMA

Divergence[a]	Relative Absorbance[b]					
HIV-1	70.0°C	72.5°C	75.0°C	77.5°C	80.0°C	
0.00 (n = 4)	1.17 ± 0.07	1.08 ± 0.04	0.56 ± 0.02	0.21 ± 0.02	≤0.01	
0.07 (n = 3)	1.28 ± 0.09	1.09 ± 0.04	0.49 ± 0.03	0.08 ± 0.01	≤0.01	
0.15 (n = 3)	1.25 ± 0.06	0.84 ± 0.06	0.34 ± 0.06	≤0.01	≤0.01	
0.19 (n = 3)	1.23 ± 0.05	0.72 ± 0.04	0.11 ± 0.02	≤0.01	≤0.01	
0.30 (n = 3)	1.14 ± 0.06	0.40 ± 0.02	0.02 ± 0.01	≤0.01	≤0.01	
HCV	69.0°C	74.0°C	80.0°C	81.5°C	85.5°C	89.0°C
0.00 (n = 4)	1.61 ± 0.01	1.52 ± 0.05	1.25 ± 0.01	0.53 ± 0.04	0.22 ± 0.01	0.13 ± 0.04
0.07 (n = 4)	1.54 ± 0.07	1.25 ± 0.04	1.15 ± 0.03	0.29 ± 0.03	0.04 ± 0.01	0.02 ± 0.01
0.26 (n = 4)	1.63 ± 0.06	0.53 ± 0.05	0.19 ± 0.03	0.07 ± 0.02	≤0.01	≤0.01
0.39 (n = 8)	1.56 ± 0.04	0.36 ± 0.04	0.24 ± 0.06	≤0.01	≤0.01	≤0.01

[a]HIV-1 *env* and HCV *C-E1* homoduplexes and heteroduplexes were derived from plasmids containing HIV-1 env and HCV C-E1 coding sequences, respectively. Numbers in parentheses indicate the number of independent capture assays.

[b] The absorbance measurements under each hybridization temperature (five for HIV-1 and six for HCV) were derived from 11 independent experiments (eleven 96-well plates) carried out at the corresponding hybridization temperature. For each microwell, the relative absorbance measurements were derived as follows: [absorbance (490 nm)] − [average absorbance from eight negative controls included in each 96-well plate]. Relative absorbance measurements are average measurements from independent capture assays (± SD).

15%; the divergence between strains from any two different subtypes (indicated by letters A–C) ranged from about 24% to 26%; the divergence between strains from any two different genotypes (indicated by numbers 1–3) ranged from about 35% to 40%. Again, depending on the total number and the diversities of HCV *C-E1* sequences analyzed by similar phylogenetic analysis, the intra- and intersubtype and intergenotype diversities may vary by about 3–5%. The left panel of Figure 1B presents the results of COMA (derived from Table 1). The four melting curves obtained for HCV define the three boundaries between hybridization signals from intrasubtype, intersubtype, and intergenotype heteroduplexes. The upper melting curve (diamond) presents hybridization signals that resulted from submitting to COMA DNA homoduplexes derived from single HCV strains (0% divergence). The next lower melting curve (square) presents hybridization signals that resulted from submitting to the assay DNA heteroduplexes derived from strains from within the same HCV subtype (7% divergence). The next lower melting curve (triangle) presents hybridization signals that resulted from submitting to the assay DNA heteroduplexes derived from strains from different HCV subtypes (26% divergence). The lowest melting curve (X) presents hybridization signals that resulted from submitting to the assay DNA heteroduplexes derived from strains from different HCV genotypes (39% divergence). The T_m for the HCV DNA homoduplex curve was 81.5 ± 0.5°C; the T_m for the intrasubtype DNA heteroduplex curve was 81 ± 0.5°C; the T_m for the intersubtype DNA heteroduplex curve was 73 ± 0.5°C; and the T_m for the intergenotype DNA heteroduplex curve was 72 ± 0.5°C. As with the absorbance ratios found in the HIV-1 studies, the aborbance ratios found with HCV reveal that the temperature range of 77° to 82°C is sufficient to distinguish intrasubtype from intersubtype heteroduplexes.

Combinatorial Melting Assay

Figure 2 schematically illustrates how COMA identified HIV-1 and HCV genetic subtypes on the basis of the results of the differential melting

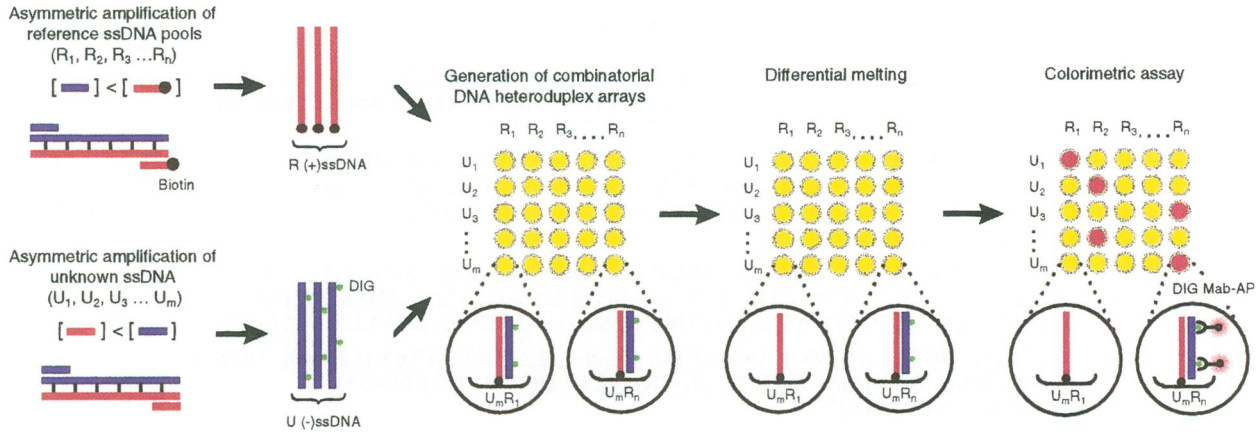

Fig. 2. Schematic illustration of the combinatorial melting assay.

curves (Fig. 1). In brief, antisense single-stranded DNA molecules were generated for each reference subtype by asymmetrical PCR. Biotinylation of the 5′ end of these reference DNAs was achieved by conjugating biotin to the appropriate primer for PCR, which in turn allowed these molecules to serve as capture DNA after binding to streptavidin-coated plates. Up to 27 reference HIV-1 DNAs (Table 2) and 6 reference HCV DNAs (Table 3) from the same viral subtype were combined to form reference pools that were then individually placed in columns of the assay plate (Fig. 3). Complementary sense-stranded molecules from unknown DNA were generated by asymmetrical PCR, during which antigenic labeling with digoxigenin was carried out. These unknown DNAs were then placed in rows of the assay plate, thus allowing the formation of heteroduplexes between each unknown DNA and each reference pool. The hybridization step was usually carried out at 75 ± 1°C for HIV-1 and 80 ± 1°C for HCV. The genetic subtype for each unknown HIV-1 or HCV sample was subsequently determined by a colorimetric immunoassay using an alkaline phosphates-conjugated monoclonal antibody directed against digoxigenin. Unknown HIV-1 and HCV samples were genotyped on the basis of whether the comparative absorbance measurements from heteroduplexes that formed with pools of the correct subtype differed from those of any other by a factor of at least two.

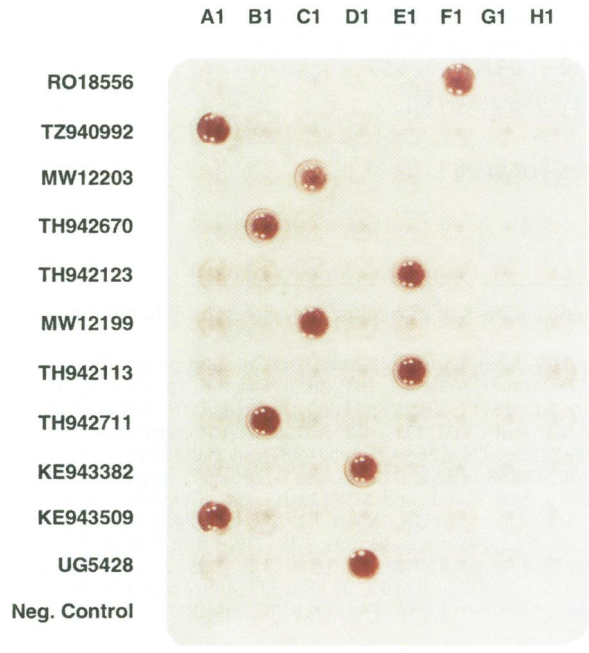

Fig. 3. An example of a final readout of the combinatorial melting assay.

HIV-1 and HCV COMA Evaluation

The sensitivity and specificity of COMA in identifying HIV-1 subtypes (A–H) were evaluated on 126 viral samples obtained from HIV-1-infected individuals from 15 different countries. The genetic subtypes of these samples had previously been determined by either nucleotide sequencing and phylogenetic analysis of *env* sequences (101 samples) or heteroduplex mobility assay (25 samples) (8–9). Table 2 provides a summary of samples containing subtypes A–H that were tested by COMA as coded unknowns. Complete agreement was obtained between the results from COMA and those from phylogenetic analysis or heteroduplex mobility assay. The relative

Table 2. Genetic subtypes of HIV-1 strains determined by COMA

Subtype[a]	Name (GenBank accession number)	Reference Pool[c]
A	UG271 (*8*), UG273 (*L22957*), UG275 (*L22951*), UG278 (*8*), DJ263 (*L22941*), UG276 (*8*), UG13655,[b] TZ940990 (*U61875*), TZ940992 (*U61878*), KE943509,[b] TX940996 (*U61876*), KE943511,[b] TZ941004 (*U61877*), KE943505[b]	I
B	POC367 (*8*), POC369 (*8*), POC370 (*8*), EG147 (*8*), POC366 (*8*), TH942705 (*U22552*), TH942728 (*U22600*), TH942706 (*U22554*), TH942664 (*U22581*), TH942698 (*U22543*), TH942711 (*U22560*), TH942670 (*U22587*), TH942699 (*U22604*), TH942723 (*U22595*)	I
C	SE145 (*L22946*), DJ259 (*L22940*), DJ260 (*8*), BR15 (*M74984*), MW6510 (*L15733*), MW12199 (*L15721*), MW12233 (*L15730*), MW12205 (*L15723*), MW6506 (*L15731*), MW12225 (*L15727*), MW6512 (*L15734*), MW6518 (*L15735*), MW12203 (*L15722*)	I
D	UG272 (*8*), UG274 (*L22950*), UG267 (*8*), UG269 (*L22949*), SE365 (*L22945*), UG279 (*8*), UG35 (*8*), UG24 (*8*), TAN372 (*8*), UG5428,[b] UG13671,[b] KE943513,[b] KE943384,[b] KE943382[b]	I
E	TH942095 (*U22611*), TH942123 (*U22557*), TH942113 (*U22625*), TH942125 (*U22561*), TH942093 (*U22609*), BZ162 (*L11751*), BZ163 (*L22946*)	I
F	BZ126 (*L22082*), RO18556,[b] RO18572,[b] RO18596,[b] RO18578[b]	I
A	KE943507[b]	II
B	TH942700 (*U22545*), TH942724 (*U22596*)	II
C	MW12227 (*L15728*)	II
E	TH942115 (*U22542*), TH942127 (*U22567*), TH942129 (*U22575*), TH942111 (*U22624*), TH942105 (*U22617*), TH942686 (*U22604*), TH942121 (*U22553*), TH942117 (*U22548*), TH942097 (*U22612*)	II
F	RO18558,[b] RO18576,[b] RO18568,[b] RO18584,[b] RO18552,[b] RO18570,[b] RO18594[b]	II
A	UG29 (*8*), UG31 (*L34667*)	III
B	BZ195 (*8*)	III
C	DJ261 (*8*), DJ262 (*8*)	III
D	TAN371 (*8*), UG277 (*8*)	III
A	UG13673,[b] UG13675,[b] UG13669[b]	IV
C	MW6508 (*L15732*), MW12215 (*L15727*), MW6514[b]	IV
D	TZ940994 (*U61879*)	IV
A	92UG037 (*U09127*), 92RW020 (*U08794*)	V
B	91US005 (*U27434*), 92HT953 (*U08444*), 92HT954 (*U08445*), 91HT651 (*U08441*), 92HT596 (*U08446*), 91HT652 (*U08443*), 92HT599 (*U08447*), 92US716 (*U08453*), 92US712 (*U08449*), 92US715 (*U08451*), 92US714 (*U08448*), 92US714 (*U08450*)	V
C	93MW960 (*U08454*), 93MW959 (*U08453*), 93MW956 (*U08455*)	V
D	92UG021 (*U27399*)	V
E	93TH975 (*U08457*), 93TH966 (*U08456*), 93TH975 (*U08458*)	V
F	93BR019 (*U27401*), 93BR020 (*U27408*)	V
G	92RU131 (*U27445*)	V

[a]HIV-1 genetic subtypes had been previously determined by phylogenetic analyses of nucleotide sequences or by heteroduplex mobility assay (8,9).

[b]Unpublished sequences communicated by M. L. Kalish. The following HIV-1 *env* sequences were used as subtype reference pools: A1: (RW20, IC144 and SF170); A2: (RW20, IC144, SF170, UG278, and DJ263); A3: (RW20, IC144, SF170, UG278, DJ263, UG273, UG275, UG13655, TZ940990, TZ940992, KE943509, TZ940996, KE943511, and TZ941004); A4: (RW20, UG271, UG273, and UG31); B1: (BR20, TH14, CYHO401, CYHO111, CYHO503, CYHO481, and CYHO271); B2: (BR20, TH14, CYHO401, CYHO111, CYHO503, CYHO481, CYHO271, TH942705, TH942728, TH942706, TH942664, TH942698, and TH942711); B3: (BR20, TH14, and SF162); C1: (MA959, IN868, and BR25); C2: (MA959, IN868, BR25, DJ259, and DJ260); C3: (MA959, IN868, BR25, DJ259, DJ260, SM145, MW6510, MW12199, MW12233, MW6512, MW12203, and MW12205); C4: (MA959, BR25, MW6508, MW12227, and MW12215); D1: (UG21, UG38, and UG46); D2: (UG21, UG38, UG46, UG272, UG274, UG279, TAN371, UG35, and UG24); D3: (UG21, UG38, UG46, UG269, UG274, UG5428, and KE943382); E1: (TH22 and TH06); E2: (TH22, TH06, and TH942105); E3: (TH22, TH06, TH942105, TH942095, TH942123, TH942113, TH942125, TH942111, TH942115, TH942127, TH942093, and TH942686); F1: (BZ126, BZ162, and BZ163); F2: (BZ126, BZ162, BZ163, RO18596, and RO18578); F3: (BZ126, BZ162, BZ163, RO18596, RO18578, RO942095, RO942123, RO18596, RO18588, RO18576, and RO18566); G1: (RU131, LBV21-7, and V1525); G2: (RU131B1); H1: (CA13 and VI557).

[c]The following subtype-reference pools were used in the combinatorial melting assay: I(A1, B1, C1, D1, E1, F1, G1, and H1); II (A2, B1, C2, D1, E2, F2, G1, and H1); III (A2, B1, C2, D2, E2, F2, G1, and H1); IV (A3, B2, C3, D3, E3, F3, G1, and H1); V (A4, B3, C4, D1, E1, F1, and G2).

Table 3. Genetic subtypes of HCV strains determined by COMA

Subtype[a]	Name
1A	USCHI0, USCHI152, USCHI168, USCHI174, USIRW09333, USIRW13505, USIRW13862, USIRW22204, USIRW22723, USIRW14626, USIRW16874, USIRW23591, USIRW20143, USIRW20317, USIRW00320, USIRW61031, USIRW61314, USIRW37125
1B	ITCHI1, ITCHI2, ITCHI24, USCHI132, FRCHIN8, USIRW18931, USIRW19585, USIRW20279, USIRW24497, USIRW19816, USIRW56176, USIRW34999, USIRW10640
2A	USCHI140, USCHI173, USCHI155
2B	USCHI162, USCHI160
2C	ITCHI23, ITCHI33, ITCHIN2
3A	ITCHI7, ITCHI20, ITCHI22, ITCHI26, ITCHI35, ITCHI164, USCHI145, USIRW14395, USIRW14974

[a]HCV genetic subtypes were determined by phylogenetic analyses of the *C-E1* nucleotide sequences. The following HCV *C-E1* sequences were used as subtype reference pools: 1A: (USCHI0, USCHI152, USCHI168, and USCHI174); 1B: (ITCHI1, ITCHI2, ITCHI24, and ITCHI132); 2A: (USCHI140, USCHI155, and USCHI173); 2B: (USCHI160 and USCHI162); 2C: (ITCHI23 and ITCHI33); 3A: (ITCHI20, ITCHI22, ITCHI26, ITCHI35, USCHI145, and USCHI164).

absorbance density ratio of "signal" to "background" was on average about 10 to 1 for 984 combinatorial measurements. In fact, in most cases the results were usually evident to the naked eye. A small portion (~5%) of HIV-1 samples initially gave ambiguous results (i.e., the absorbance density from heteroduplexes that formed with pools of the correct subtype did not differ from those of any other by a factor of at least two). As a possible explanation for these ambiguities, it is possible that an inadequate variety of known strains within the correct reference pool prevented the unknown strain from forming a heteroduplex of sufficiently high complementarity, compared with the heteroduplexes from the other pools, at that stringent temperature. To resolve the ambiguous results, we increased either the number of reference DNA in the relevant pools and/or the hybridization temperature (Table 2).

The utility of COMA to identify HCV types and subtypes was tested in a similar manner. Forty-eight HCV samples were first genotyped by nucleotide sequencing and phylogenetic analysis (Table 3; GenBank accession number pending). These samples were then coded for testing by COMA, which again yielded genotyping results that completely agreed with those obtained by DNA sequencing (Table 4). Unlike HIV-1, all the unknown HCV samples gave unambiguous results with a limited number of reference DNA in each reference pool because the topology of HCV phylogenetic tree is different from that of HIV-1. The average intrasubtype diversity of the HCV *C-E1* region is less than the corresponding intrasubtype diversity of the HIV-1 *env* (see Fig. 1).

Discussion

The development of COMA draws from the experiences of numerous investigators over many years in the area of DNA melting kinetics (11,12), particularly with regard to work on denaturing gradient gel electrophoresis (13–17). Many techniques have been developed to measure the melting and the reannealing processes of duplex DNA molecules, but critical questions about the kinetics of these processes remain unanswered. COMA is a simple comparative technique that uses combinatorial solution-phase DNA hybridization measurements to determine the genetic homologies between related DNA sequences. To genotype unknown HIV-1 and HCV strains, we used several collections of reference single-stranded DNA molecules representative of each genetic subtype or genotype (for HCV). The simple rationale of COMA is that, at "stringent" hybridization temperatures, the degree of complementary of the unknown single-stranded DNA molecule to complementary DNA molecules from the same genetic subtype is higher than that from other subtypes. This principle of operation offers a major advantage to COMA compared with existing genotyping techniques utilizing analytical gel eletrophoresis: the abil-

Table 4. Comparative absorbance ratios obtained by COMA and determination of the genetic subtypes of HIV-1 strains

Name (GenBank accession number)	Subtype	Country Origin	Comparative Absorbance Ratios								Subtype Determined by COMA
			A	B	C	D	E	F	G	H	
RO18566[a]	F	Romania	0.08	0.09	0.02	0.01	0.01	**1.00**	0.00	0.02	F
TZ940992 (*U61878*)	A	Tanzania	**1.00**	0.12	0.08	0.09	0.06	0.09	0.03	0.08	A
MW12203 (*L15722*)	C	Malawi	0.00	0.00	**1.00**	0.00	0.00	0.00	0.01	0.00	C
TH942670 (*U22587*)	B	Thailand	0.01	**1.00**	0.03	0.00	0.00	0.00	0.00	0.00	B
TH942123 (*U22557*)	E	Thailand	0.11	0.09	0.05	0.03	**1.00**	0.03	0.01	0.03	E
MW12199 (*L15721*)	C	Malawi	0.14	0.07	**1.00**	0.08	0.15	0.07	0.04	0.10	C
TH942113 (*U22625*)	E	Thailand	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.31	E
TH942711 (*U22560*)	B	Thailand	0.04	**1.00**	0.03	0.01	0.01	0.00	0.00	0.00	B
KE943382[a]	D	Kenya	0.02	0.00	0.01	**1.00**	0.00	0.00	0.00	0.00	D
KE943509[a]	A	Kenya	**1.00**	0.16	0.07	0.09	0.07	0.14	0.03	0.06	A
UG5428[a]	D	Uganda	0.09	0.03	0.01	**1.00**	0.00	0.01	0.00	0.00	D

The comparative absorbance ratios were determined as explained in Materials and Methods using the subtype-reference pools described in Table 2. HIV-1 genetic subtypes had been previously determined by phylogenetic analyses of nucleotide sequences.
[a]Unpublished sequences communicated by M. L. Kalish. For each unknown sample, the reference pool which produces the highest relative absorbance (comparative absorbance ration = 1.00), shown in bold numbers, indicates the correct genetic subtype. Hybridization temperature was 76 ± 0.5°C.

ity to compare each unknown HIV-1 and HCV strain with multiple reference strains simultaneously.

The extensive genetic variability of HIV-1 and HCV has important implications for future vaccine development against these viruses, as well as for the clinical management of HCV-infected patients (18,19). In this report we have shown that a newly developed assay provides a rapid, accurate, and simple method for the determination of HIV-1 and HCV subtypes worldwide. It should be mentioned that for assay optimization these studies represent only the initial step. Although the assay works efficiently and with high predictive value, improvements are still to be made. Nevertheless, the current assay should accelerate current efforts to understand the global molecular epidemiology of HIV-1 and HCV, especially in developing countries. given that more than half of all full-length HIV-1 genomes in the database are recombinant (20,21), the combinatorial melting assay may be a valuable tool to screen multiple genomic regions of HIV for recombinant strains.

Acknowledgments

The authors gratefully acknowledge the generosity of M. Busch, E. L. Delwart, L. Detmer, B. Hahn, F. Gao, M. L. Kalish, J. Kolberg, L. Tobler, M. Urdea, A. Weiner, and the AIDS Research and Reference Reagent Program for providing HIV-1 and HCV samples for this study and Leslie Mayer and Young Guo for technical assistance and P. G. McHardy and J. P. Moore for helpful discussions. The authors acknowledge support from the Aaron Diamond Foundation and the NIH. Leondios G. Kostrikis is an Aaron Diamond Fellow.

References

1. Myers G, Foley B, Mellors JW, Korber B, Jeang K-T, Wain-Hobson S. (1996) Human Retroviruses and AIDS. Los Alamos National Laboratory.
2. Bukh J, Purcell RH, Miller RH. (1992) Sequence analysis of the 5′ noncoding region of hepatitis C virus. *Proc. Natl. Acad. Sci. U.S.A.* **89:** 4942–4946.
3. Bukh J, Purcell RH, Miller RH. (1994) Sequence analysis of the core gene of 14 hepatitis C virus genotypes. *Proc. Natl. Acad. Sci. U.S.A.* **91:** 8239–8243.

4. Bukh J, Purcell RH, Miller RH. (1993) At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. *Proc. Natl. Acad. Sci. U.S.A.* **90:** 8234–8238.
5. Bukh J, Miller RH, Purcell RH. (1995) Genetic heterogeneity of hepatitis C virus: Quasispecies and genotypes. *Semin. Liver Dis.* **15:** 41–63.
6. Simmonds P, Holmes EC, Cha TA, et al. (1993) Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J. Gen. Virol.* **74:** 2391–2399.
7. Smith DB, Pantirana S, Davidson F, et al. (1997) The origin of hepatitis C virus genotypes. *J. Gen. Virol.* **78:** 321–328.
8. Delwart EL, Shpaer EG, Louwagie J, et al. (1993) Genetic relationships determined by a DNA heteroduplex mobility assay: Analysis of HIV-1 *env* genes. *Science* **262:** 1257–1261.
9. Gao F, Morrison SG, Robertson DL, et al. for WHO and NIAID Networks for HIV Isolation and Characterization. (1996) Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. *J. Virol.* **70:** 1651–1667.
10. Kostrikis LG, Bagdades E, Cao Y, Zhang L, Dimitriou D, Ho DD. (1995) Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: Identification of a new subtype designated subtype I. *J. Virol.* **69:** 6122–6123.
11. Wetmur JG, Davidson N. (1968) Kinetics of renaturation of DNA. *J. Mol. Biol.* **31:** 349–370.
12. Bonner TI, Brenner DJ, Neufeld BR, Britten RJ. (1973) Reduction in the rate of DNA reassociation by sequence divergence. *J. Mol. Biol.* **81:** 123–135.
13. Andersson B, Ying JH, Lewis DE, Gibbs RA. (1993) Rapid characterization of HIV-1 sequence diversity using denaturing gradient gel electrophoresis and direct automated DNA sequencing of PCR products. *PCR Methods Appl.* **2:** 293–300.
14. Myers RM, Fischer SG, Maniatis T, Lerman LS. (1985) Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis. *Nucl. Acids Res.* **13:** 3111–3129.
15. Myers RM, Maniatis T, Lerman LS. (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* **155:** 501–527.
16. Myers RM, Lumelsky N, Lerman LS, Maniatis T. (1985) Detection of single base substitutions in total genomic DNA. *Nature* **313:** 495–498.
17. Myers RM, Fischer SG, Lerman LS, Maniatis T. (1985) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucl. Acids Res.* **13:** 3131–3145.
18. Sharara AI, Hunt CM, Hamilton JD. (1996) Hepatitis C. *Ann. Intern. Med.* **125:** 658–668.
19. Zein NN, Rakela J, Krawitt EL, Reddy KR, Tominaga T, Persing DH, Collaborative Study Group. (1996) Hepatitis C virus genotypes in the United States: Epidemiology, pathogenicity, and response to interferon therapy. *Ann. Intern. Med.* **125:** 634–639.
20. Robertson DL, Hahn BH, Sharp PM. (1995) Recombination in AIDS viruses. *J. Mol. Evol.* **40:** 249–259.
21. Robertson DL, Sharp PM, McCutchan FE, Hahn BH. (1995) Recombination in HIV-1. *Nature* **374:** 124–126.