An African primate lentivirus (SIV_{sm}) closely related to HIV-2

Vanessa M. Hirsch*, Robert A. Olmsted*, Michael Murphey-Corb†, Robert H. Purcell‡ & Philip R. Johnson*

* Division of Molecular Virology and Immunology, Department of Microbiology, Georgetown University, NIH/Twinbrook II, 12441 Parklawn Drive, Rockville, Maryland 20852, USA
† Delta Regional Primate Research Center, Tulane University, Covington, Maryland Louisiana 70433, USA
‡ Laboratory of Infectious Diseases, NIAID, NIH, Bethesda, Maryland 20852, USA

THE ancestors of the human immunodeficiency viruses (HIV-1 and HIV-2) may have evolved from a reservoir of African non-human primate lentiviruses, termed simian immunodeficiency viruses (SIV)[1]. None of the SIV strains characterized so far are closely related to HIV-1[2–6]. HIV-2, however, is closely related to SIV (SIV_{mac}) isolated from captive rhesus macaques (*Macaca mulatta*)[7]. SIV infection of feral Asian macaques has not been demonstrated by serological surveys[8,9]. Thus, macaques may have acquired SIV in captivity by cross-species transmission from an SIV-infected African primate. Sooty mangabeys (*Cercocebus atys*), an African primate species indigenous to West Africa, however, are infected with SIV (SIV_{sm}) both in captivity[9–11] and in the wild (P. Fultz, personal communication). We have molecularly cloned and sequenced SIV_{sm} and report here that it is closely related to SIV_{mac} and HIV-2. These results suggest that SIV_{sm} has infected macaques in captivity and humans in West Africa and evolved as SIV_{mac} and HIV-2, respectively.

We have previously described the derivation of five genome-length proviral DNA clones of $SIV_{sm/Delta/F236}$ (ref. 12). This virus was isolated in 1986 at the Delta Regional Primate Research Center from a rhesus macaque (F236) that died from opportunistic infections after experimental inoculation with SIV isolated from an asymptomatic sooty mangabey (E038). Mangabey E038 had been obtained from the Yerkes Regional Primate Research Center[13]. On transfection into H9 cells in culture, two of the five clones (λsmH-3 and λsmH-4) produced infectious virions. Progeny of both biologically active molecular clones established persistent infections in experimentally inoculated rhesus and pig-tail macaques (data not shown).

The complete nucleotide sequence (10,241 base pairs (bp)) of the infectious proviral molecular clone λsmH-4 was determined and analysed. The genomic organization of SIV_{sm} was similar to SIV_{mac} and HIV-2 (Fig. 1). The long terminal repeat (LTR) of SIV_{sm} (815 bp) was 87% identical to SIV_{mac} and 74% to HIV-2 overall. Regulatory sequences, including potential Sp1-binding regions and the TAR sequence, were highly conserved. Comparison of the deduced amino-acid sequences from the nine open reading frames in the SIV_{sm} genome revealed significant similarities with SIV_{mac} and HIV-2 (Table 1). The regions encoding Gag (Fig. 2) and Pol proteins were highly conserved, accounting for the serological cross-reactivity previously noted among the three viruses[1]. An interesting feature of the λsmH-4 *pol* nucleotide sequence was a single-base

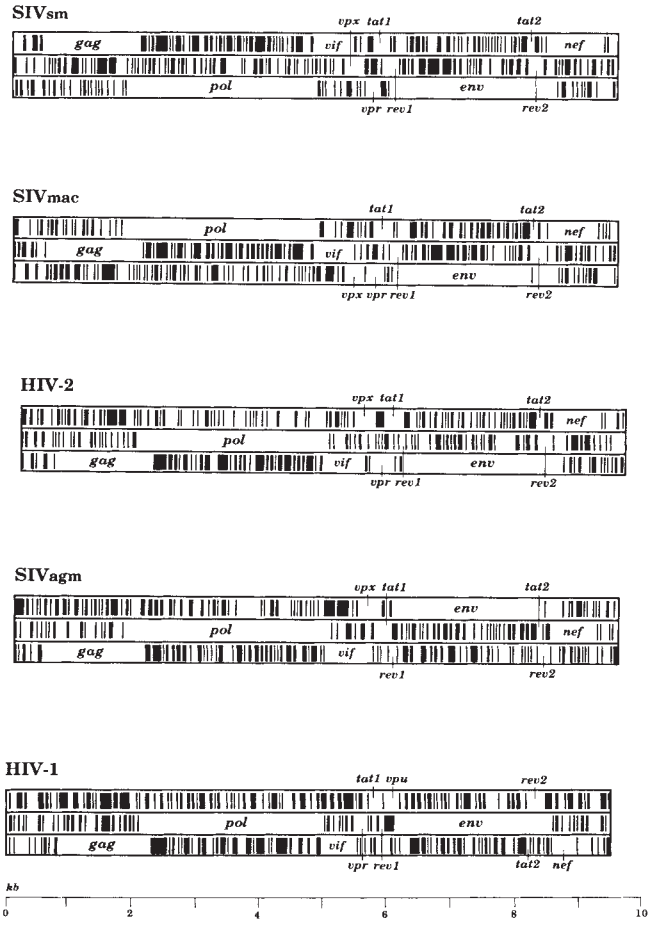

FIG. 1 Schematic representation of the open reading frames of five primate lentiviruses. Bars in each frame represent positions of stop codons. The sequences displayed are: SIV_{sm} (strain F236); SIV_{mac} (strain 142); HIV-2 (strain ROD); SIV_{agm} (strain TYO-1); and HIV-1 (HXB2).

TABLE 1 Amino-acid identity (%) between proteins of primate lentiviruses

								Envelope					
										gp32/41			
	Gag	Pol	Vif	Vpr	Rev	Tat	Vpx	gp160	gp120	Whole	Before stop codon	After stop codon	Nef
$SIV_{sm/Delta/F236}$ versus:													
SIV_{mac}142	90	92	84	90	70	72	95	81	81	81	89	70	77
HIV-2 (ROD)	84	84	69	70	62	65	80	72	70	76	87	66	64
SIV_{agm} (TYO-1)	59	58	32	—	43	32	38	46	45	47	56	36	35
HIV-1 (HXB-2)	55	57	28	45	32	26	—	36	34	41	43	37	36
SIV_{mac}: 142 versus 251	96	97	90	98	88	88	97	92	91	94	94	93	82
HIV-2: ROD versus NIH-Z	92	91	86	85	86	85	84	80	78	84	89	78	—
SIV_{mac} 142 versus HIV-2 ROD	82	81	64	70	59	63	86	70	68	73	87	64	73

Sequences obtained from the Los Alamos HIV Sequence Database. Alignments performed using the PRTALN program[23] with default parameters. Amino-acid identities calculated with the inclusion of insertions and deletions. Dashed line indicates no comparison performed because: SIV_{agm} lacks vpr; HIV-1 lacks vpx; or because of a large deletion in Nef of NIH-Z.

insertion near the putative proteolytic cleavage site for the integrase protein. This insertion introduced a frameshift and subsequent stop codon in the extreme 5′ end of the integrase nucleotide sequence. The predicted amino-acid sequence of λsmH-3 in this region was analysed and found to be identical to SIV_{mac} and HIV-2. It was interesting to note that after transfection of λsmH-4 into H9 cells in culture, reverse transcriptase activity was detected at 30 days, as opposed to 7 days for λsmH-3 (ref. 12). Experiments are in progress to define further the biological importance of this single-base insertion.

Genes in the complex central region were variably conserved (Table 1). Surprisingly, the essential regulatory genes, *rev* and *tat*, were relatively poorly conserved. However, distinctive structural motifs within *rev* and *tat* (for example, the arginine-rich domain in the second coding exon in *rev* and the cysteine cluster in the first coding exon of *tat*) were maintained among SIV_{sm}, SIV_{mac} and HIV-2. The most highly conserved central region gene is *vpx*, an open reading frame common to all characterized strains of SIV and HIV-2. The *vpx* gene product ($p14^{vpx}$) is immunogenic in natural HIV-2 and SIV infection[14,15] and may bind to the single-stranded RNA genome in mature virions[16].

Alignment of the predicted envelope-protein sequences of SIV_{sm}, SIV_{mac} and HIV-2 reveals an overall conservation of cysteine residues and other primary structural features (Fig. 3). Variable domains in *env*, predicted by analogy to HIV-1, were also variable among SIV_{sm}, SIV_{mac} and HIV-2. In addition, another highly variable region was identified within the putative cytoplasmic domain of the gp40 transmembrane subunit (residues 736–752). This area contains an in-frame stop codon that markedly truncates the cytoplasmic domain in some SIV_{mac} and HIV-2 molecular clones[2–4,17]. The SIV_{sm} *env* nucleotide sequence does not contain a premature stop-codon and is predicted to encode a full-length gp40, in contrast to the gp32 of SIV_{mac} and several HIV-2 isolates[18–20]. This prediction was confirmed by the demonstration of a transmembrane protein of relative molecular mass (M_r) 40,000 on western blots of virions collected after transfection of the λsmH-4 and λsmH-3 clones into H9 cells[12]. The reduction in sequence similarity among the three envelope proteins after the stop codon region was significant (Table 1; Fig. 2). This may indicate either a lack of structural (or functional) constraints on the amino-acid sequence of the cytoplasmic domain or variable suppression or mutation of the stop codon resulting in a mixed population of viruses, or both. The full-length cytoplasmic domain seems to be synthesized during SIV_{mac} infection of macaques, but is not detected in SIV_{mac}-infected cells in tissue culture (ref. 21; M.M.-C, unpublished data). Thus, one possibility is that SIV_{mac} and HIV-2 molecular clones containing the premature stop-

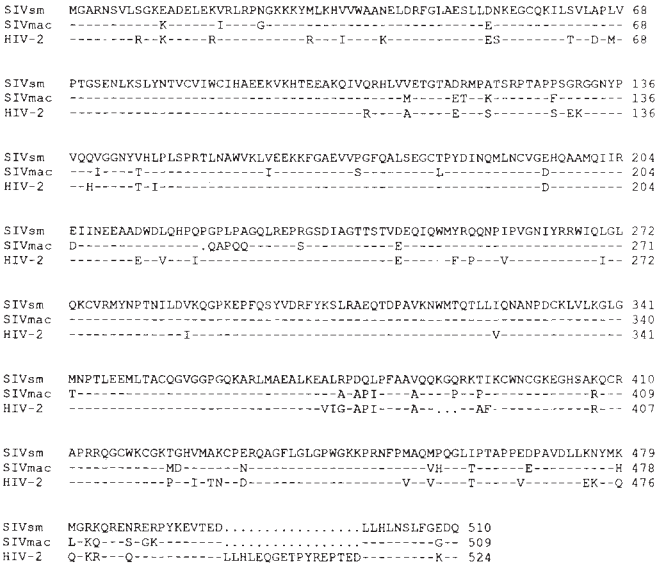

FIG. 2 Alignment of the predicted Gag protein sequences of $SIV_{sm/Delta/F236}$, SIV_{mac} (251) and HIV-2 (ROD). Dashes below the SIV_{sm} sequence, identical amino acids; dots, gaps introduced to maximize the alignment.
METHODS. λsmH-4 was subcloned (in 1–3-kb fragments) into a plasmid vector (pTZ18R; Pharmacia) and sequenced by the dideoxynucleotide chain-terminating method, using T7 DNA polymerase (United States Biochemical). The *gag* nucleotide sequence was derived from subclone psmH-4B including the 5′ non-coding *gag* sequence and a 5′ portion of the *pol* gene. DNA sequences were analysed using the PC/Gene, IFIND, SEQ and GenALIGN programs (IntelliGenetics).

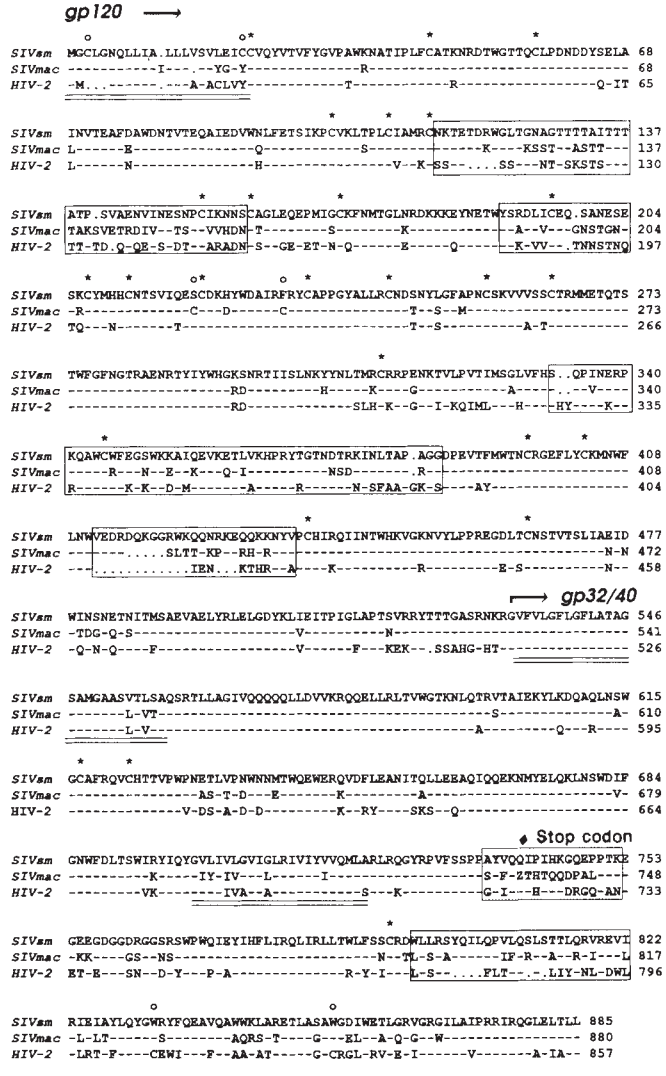

FIG. 3 Alignment of the predicted envelope glycoproteins of $SIV_{sm/Delta/F236}$, SIV_{mac} (251) and HIV-2 (ROD). Conventions for the alignment as in Fig. 1. Boxes, variable domains (as defined among HIV-1 isolates). In addition, another variable domain in gp40 surrounding the premature stop codon in SIV_{mac} (black diamond) is enclosed by a shaded box. Cysteines conserved among all three sequences indicated by asterisks; non-conserved cysteines marked by open circles above the sequence. The putative proteolytic cleavage site between gp120 and gp40 is marked by an arrow. Predicted hydrophobic domains (signal, N terminus of gp40, and the membrane-spanning domain) are indicated by double underlining.
METHODS. The nucleotide sequence of the SIV_{sm} *env* gene was determined (as described in Fig. 1) from three plasmid subclones of λsmH-4: psmH-4C2 (gp120); psmH-4D (gp40); and, psmH-4E (C terminus of gp40). DNA sequence analysis and alignments performed as in Fig. 1.

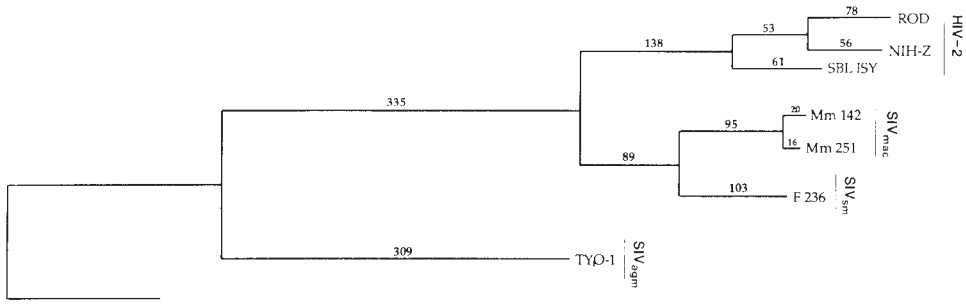

FIG. 4 Minimum-length evolutionary tree based on *gag* nucleotide sequences of the SIV/HIV-2 lentivirus subfamily.
METHODS. The tree was constructed from variation in the *gag* nucleotide sequence using PAUP (ref. 26) with the global branch-swapping option MULPARB. The total number of sites examined was 1,748, of which 858 were variable. The minimum length of the tree was 1,690 and the consistency index was 0.848. The tree was rooted on a random sequence (as in ref. 22). Lengths of the horizontal lines are proportional to the minimum number of single nucleotide substitutions required to generate the observed variation (also indicated by the numeric branch lengths above each line). The length of the vertical lines is for clarity only. Evolutionary tree analyses based on deduced protein sequences[27] generated a similar branching order of divergence.

codon represent viruses artificially selected for their ability to grow in tissue culture. We are investigating the functional importance of this highly variable region.

A distinctive feature of the SIV_{sm} sequence is an extended open reading frame for *nef*, the negative regulatory gene. In HIV-1, HIV-2 and SIV_{mac} molecular clones, the *nef* gene is variable in length; in general, the HIV-1 *nef* gene is truncated relative to SIV and HIV-2. The SIV_{sm} *nef* gene contains an apparent frameshift (single-base insertion) near the C terminus that extends the predicted reading frame for an additional 54 amino acids relative to SIV_{mac} and HIV-2. The sequences of both SIV_{sm} biologically active clones are identical through this region, indicating that the extended reading frame is not limited to a single clone.

To investigate further the relationship of SIV_{sm} to SIV_{mac} and HIV-2, we constructed minimum-length evolutionary trees[22] based on the nucleotide variations in the LTR and the *gag* and *env* genes (the *gag* nucleotide tree is shown in Fig. 4). The trees are all consistent in the branching order of divergence, although branch lengths vary (*env* genes are more divergent). By this analysis of *gag*, SIV_{sm} was at least 12% different from SIV_{mac} and 26–28% different from HIV-2. A precise time calibration for this tree was not possible because the mutation rates for SIV_{mac} and HIV-2 have not been determined and could not be assumed to be equal or constant. However, if we assume a representative mutation rate for the *gag* gene of 0.5% per virus per year (estimated from HIV-1 (ref. 22), SIV_{sm} and SIV_{mac} might have diverged from each other 10 years ago (minimum), and SIV_{sm}/SIV_{mac} might have diverged from HIV-2 as recently as 30 years ago. The latter figure is in agreement with previously published data[23]. We emphasize that these calculations are only rough estimates and are based on mutation rates that have not been experimentally determined for SIV or HIV-2.

These evolutionary estimates are consistent with the limited information regarding the origins of SIV_{sm}, SIV_{mac} and HIV-2. First, consider the relationship of SIV_{sm} and SIV_{mac}. Because Asian macaques are not infected with SIV in the wild, captive macaques probably acquired SIV by cross-species transmission from an SIV-infected African primate. SIV infection of captive macaques has been traced back to the early 1970s at the New England Regional Primate Research Center; the putative index cases were macaques obtained from another primate facility (N. Letvin, personal communication). Although not documented, these macaques may have been housed in captivity with African primates (sooty mangabeys or others) before arrival in New England. If SIV_{mac} 251 (the published prototype SIV_{mac}) is thought of as a 1982 virus (the year that rhesus macaque 251 died)[24], then we estimate that SIV_{sm} and SIV_{mac} might have diverged in the late 1960s or early 1970s (based on the minimum divergence time of 10 years, estimated above). In support of this hypothesis, SIV_{sm} has probably been in the United States since before 1968, the last year that wild-caught West African mangabeys were added to the Yerkes colony[11]. Therefore, the molecular and epidemiological data together suggest that an SIV-infected sooty mangabey (or closely related species) was the source of SIV_{mac}.

The unusually close relationship of SIV from captive macaques with the human lentivirus HIV-2 has remained an enigma. We propose that our data on the genetic relatedness of SIV_{sm}, SIV_{mac} and HIV-2 provide the previously missing link. The relationship between HIV-2 and SIV_{sm} can be summarized by three observations: first, SIV-infected sooty mangabeys inhabit West Africa; second, HIV-2 infection of humans is endemic in West Africa; and, third, the genomes of SIV_{sm} and HIV-2 are closely related. Clearly, the SIV_{sm} that we have cloned and sequenced is not the direct ancestor of HIV-2 or SIV_{mac}. Our sequence represents an SIV_{sm} that has probably been in North America for more than 20 years. Also, the HIV-2 sequences reported to date probably represent 20 or more years of evolution of the virus in humans in West Africa. Therefore, the degree of sequence similarity we report here is remarkable and argues that these viruses evolved fairly recently from a common ancestor. A plausible interpretation of these data is that in the past 30–40 years SIV from a West African sooty mangabey (or closely related species) successfully infected a human and evolved as HIV-2. Our data cannot exclude the possibility that HIV-2 from a human was passed to a sooty mangabey and subsequently evolved as SIV_{sm}. Sequences of older HIV-2 and SIV isolates (from mangabeys and other species) are required to resolve these issues. Characterization of novel SIVs from primates in different regions of Africa may provide further clues to the origins of human lentiviruses. □

Received 27 February; accepted 18 April 1989.

1. Schneider, J. & Hunsmann, G. *AIDS* **2,** 1–9 (1988).
2. Chakrabarti, L. *et al. Nature* **328,** 543–547 (1987).
3. Franchini, G. *et al. Nature* **328,** 538–543 (1987).
4. Hirsch, V., Reidel, N. & Mullins, J. I. *Cell* **49,** 307–319 (1987).
5. Fukasawa, M. *et al. Nature* **333,** 457–461 (1988).
6. Tsujimoto, H. *et al. J. Virol.* **62,** 4044–4055 (1988).
7. Daniel, M. D. *et al. Science* **228,** 1201–1204 (1985).
8. Ohta, Y. *et al. Int. J. Cancer* **41,** 115–122 (1988).
9. Lowenstine, L. J. *et al. Int. J. Cancer* **38,** 563–574 (1986).
10. Murphey-Corb, M. *et al. Nature* **321,** 435–437 (1986).
11. Fultz, P. N. *et al. Proc. natn. Acad. Sci. U.S.A.* **83,** 5286–5290 (1986).
12. Hirsch, V. M. *et al. J. med. Primatol.* (in the press).
13. Baskin, G. B., Murphey-Corb, M., Watson, E. A. & Martin, L. N. *Vet. Path.* **25,** 456–467 (1988).
14. Yu, X-F., Ito, S., Essex, M. & Lee, T-H. *Nature* **335,** 262–265 (1988).
15. Franchini, G., Ruschie, J., O'Keeffe, T. & Wong-Staal, F. *AIDS Res. Hum. Retroviruses* **4,** 243–259 (1988).
16. Henderson, L. E., Sowder, R. C., Copeland, T. D., Benveniste, R. E. & Oroszlan, S. *Science* **241,** 199–201 (1988).
17. Guyader, M. *et al. Nature* **326,** 662–669 (1985).
18. Barin, F. *et al. Lancet* II, 1387–1389 (1985).
19. Albert, J. *et al. AIDS Res. Hum. Retroviruses* **3,** 3–10 (1987).
20. Clavel, F. *et al. Science* **233,** 343–346 (1986).
21. Franchini, G., Kanaki, P. J., Bosch, M., Fargnoli, K. & Wong-Staal, F. *AIDS Res. Hum. Retroviruses* **4,** 251–258 (1988).
22. Smith, T. F., Srinivasan, A., Schochetman, G., Marcus, M. & Myers, G. *Nature* **333,** 573–575 (1988).
23. Sharp, P. M. & Li, W.-H. *Nature* **336,** 351 (1988).
24. Hunt, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **80,** 5085–5089 (1983).

25. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80,** 726–730 (1983).
26. Day, W. H. E. *J. theor. Biol.* **103,** 429–438 (1983).
27. Feng, E. F. & Doolittle, R. *J. molec. Evol.* **25,** 351–360 (1987).

ACKNOWLEDGEMENTS. The authors thank Drs G. Myers and R. Doolittle for evolutionary tree analyses and discussions, G. Dapolito, C. McGann and S. Kitov for technical assistance, and Dr R. Chanock for discussions and support. The nucleotide sequence described in this paper has been deposited in the GenBank/EMBL database under the accession number X14307. The entire sequence can be obtained through GenBank/EMBL or by writing to the authors.

Binding of labelled influenza matrix peptide to HLA DR in living B lymphoid cells

Ruggero Ceppellini*, Guido Frumento*, Giovanni B. Ferrara*†, Roberto Tosi‡, Alberto Chersi§ & Benvenuto Pernis‖

* Istituto Nazionale per la Ricerca sul Cancro, Genova, Italy
† Cattedra di Genetica Umana, Universita di Napoli and Centro Ricerche Immunoematologiche, Avis, Bergamo, Italy
‡ Istituto di Biologia Cellulare, CNR, Roma, Italy
§ Ripartizione di Oncologia Sperimentale, Istituto Regina Elena, Roma, Italy
‖ Departments of Microbiology and Medicine, Columbia University, New York, New York 10032, USA

T CELLS recognize protein antigens as fragments (peptides) held in a defined binding site[1] of class I or class II major histocompatibility (MHC) molecules. The formation of complexes between various immunologically active peptides and different MHC molecules has been demonstrated directly in binding studies between the peptides and solubilized, purified molecules of class II MHC[2–6]. Studies with intact cells, living or fixed, have not directly demonstrated the binding of the peptides to MHC molecules on antigen-presenting cells, but the formation of such complexes has been shown indirectly through the capacity of antigen-presenting cells to stimulate specific T cells[7–9]. Here we report evidence that supports directly the binding of radiolabelled influenza matrix peptide 17–29 to products of the human class II MHC locus HLA-DR, on living homozygous B-cell lines, and we show that the kinetics of such binding is much faster with living cells than with fixed cells. Furthermore, whereas the peptide reacts with HLA-DR molecules of all alleles, it binds preferentially to DR1, the restricting element in antigen presentation.

The results of direct binding experiments with ^{125}I-labelled peptide 17-29 of influenza virus matrix protein, are reported in Fig. 1. These concern 21 HLA-homozygous Epstein-Barr virus-transformed B-cell lines from the tenth Histocompatibility Workshop, plus two human resting polyclonal T-cell populations that do not express class II MHC, and two human T-cell lines, one of which (MOLT 4) does not express class II MHC, and the other (HUT 78) expresses class II (DR 4, DRw 53, DQw 3) at a level comparable to that of the B-cell lines, as shown by fluorescence-activated cell sorter analysis with monoclonal antibody Q5/13. The MHC types of the HLA homozygous B-cell lines are given in Table 1.

The data shown in Fig. 1 indicate that peptide 17–29 of influenza matrix binds within 30 min and in reproducible amounts to living human B cells. Cells that carry the *DR1* allele, the restricting element for presentation of peptide 17–29 to T cells[10,11], show increased binding. All of our DR1-expressing cell lines also carried *DQ5*, which is in linkage disequilibrium with *DR1*, but we had one DQ5-expressing line (SP00 10) that did not express DR1, and this was not in the high-binding group. The high-binding group did however include one line that does not carry DR1—the WDV line expressing DR w13 (note that not all lines expressing Dr w13 have this property, see line WT47). All B-cell lines bound much more peptide than T cells, with an intermediate value given by HUT 78, a T-cell line that expresses class II MHC. The data shown in Fig. 1 are compatible with the possibility that peptide 17–29 of influenza matrix binds to all DR molecules, with preference for its restricting allele product, DR1.

More compelling evidence for this possibility is given by the data summarized in Fig. 2. These show that the binding of the peptide, both to a high-binder and to a low-binder B-cell line, is inhibited by an antibody for class II MHC and is not affected by an anti-class I. We obtained similar results (data not shown) with two different monoclonal anti-class I and anti-class

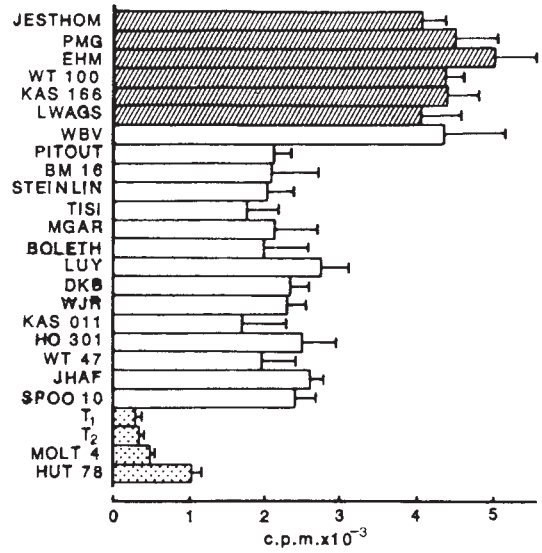

FIG. 1 Binding of influenza matrix peptide 17–29-Tyr ^{125}I to living human B and T cells. The data are the mean c.p.m. ± 1 s.d., from 18 separate experiments for EHM cells, 9 experiments for JHAF cells, 6 for BM 16 and MOLT 4, and triplicate for the other cell lines or populations. ▨, DR-1-positive B-cell lines; □, DR1-negative B-cell lines, ▩, T-cell lines (MOLT 4 and HUT 78) or polyclonal T-cell populations. Resting T cells (T_1 and T_2) were derived from the blood of two healthy donors by the E-rosette procedure.
METHODS. Peptide 17-29 of the influenza virus matrix protein (Ser-Gly-Pro-Leu-Lys-Ala-Glu-Ile-Ala-Gln-Arg-Leu-Glu) was synthesized by the solid-phase method[19] using a Du Pont-Vega Coupler, Model 100. An additional tyrosine was added at the C-terminal end of the peptide to allow iodination. *N*-t-butoxycarbonyl protection was used, and trifunctional amino acids were protected as follows: Ser (*O*-benzyl); Lys (2-chloro-benzyl-oxy-carbonyl); Glu (*O*-benzyl); Arg (tosyl); Tyr (2, 6-dichloro-benzyl). After deprotection and cleavage from the support by treatment with trifluoromethansulphonic acid) (40 min at 40 °C) (ref. 20), the free peptide was purified by gel filtration on a 2 × 100 cm column of Sephadex G25 (superfine) in 0.01 M ammonia. Thin-layer chromatography showed that the peptide was substantially homogenous, and amino-acid analysis indicated that the composition was within 5 per cent of the theoretical value. The peptide was iodinated by the chloramine-T method[21] to a specific activity of 50 μc μg^{-1} in the presence of an equimolar concentration of cold KI. Incorporation was 70–80%. Free ^{125}I was removed by gel filtration on Sephadex G-10. The labelled peptide was kept at 4 °C in 0.075 M Tris, pH 7.8, containing 2% ethanol and 0.001% sodium azide. To achieve the working concentration the peptide was diluted in complete medium immediately before testing. Experiments were performed within 3 weeks from the labelling. Binding test: Cells were washed once in RPMI 1640 and resuspended in the same medium supplemented with nutridoma HU (Boehringer), 2 mM L-glutamine and 1 mM pyruvate at a concentration of 2 × 10^7 cells per ml. Cell suspension (50 μl) was added to an equal volume of peptide solution, containing 10 ng labelled peptide in 15 ml centrifuge tubes. If not otherwise indicated, after incubation at 37 °C in CO_2 incubator for 30 min, the cells were washed three times at 4 °C and the pellet was counted in a Packard 5110 gamma counter to assay the cell-associated radioactivity.

‖ To whom correspondence should be addressed.