ARTICLES-

Genome organization and transactivation of the human immunodeficiency virus type 2

Mireille Guyader, Michael Emerman, Pierre Sonigo, François Clavel^{*}, Luc Montagnier & Marc Alizon[†]

Unité d'Oncologie Virale (CNRS UA 1157), and Unité de Recombinaison et Expression Génétique (INSERM U163, CNRS UA271), Laboratoire de Biologie Moléculaire et Immunologie des Rétrovirus, Institut Pasteur, 25 rue du Dr Roux 75724 Paris Cedex 15, France

Analysis of the nucleotide sequence of the human retrovirus associated with AIDS in West Africa, HIV-2, shows that it is evolutionarily distant from the previously characterized HIV-1. We suggest that these viruses existed long before the current AIDS epidemics. Their biological properties are conserved in spite of limited sequence homology; this may help the determination of the structure-function relationships of the different viral elements.

THE acquired immune deficiency syndrome (AIDS) has now spread worldwide and appears to be an acute public health problem in Africa in particular¹⁻⁵. A retrovirus designated human immunodeficiency virus (HIV), but previously known as LAV, HTLV-III or ARV, was shown to cause AIDS in the different areas afflicted by the epidemics⁶⁻⁸. Indeed, isolates from North America, Western Europe and Central Africa have the same biological properties, and antigenically cross-reactive proteins with the same relative molecular mass⁹⁻¹¹. Only studies at the molecular level have revealed some differences in the nucleotide sequence of North-American and African isolates^{12,13}. This sequence variation is also present, though to a lesser extent, among different isolates from the USA¹⁴⁻¹⁸.

The western part of Africa seemed relatively spared by AIDS³. Recently, however, several typical cases were found in a survey of patients from Guinea Bissau and other countries of West Africa¹⁹⁻²¹. Unexpectedly, most of these patients did not have detectable titres of antibodies against HIV. But they were found to be infected by a retrovirus related to HIV by its ultrastructural and biological properties, such as cytopathogenicity and tropism for cells carrying the CD4(T4) antigen¹⁹. Antibodies raised against HIV could immunoprecipitate the gag and pol products of these isolates, which have molecular masses that are similar but not identical to these antigens of HIV; in contrast, the env products could not be immunoprecipitated, whereas previous HIV isolates showed wide cross-antigenicity of the envelope glycoprotein. Furthermore, the genome of this new retrovirus cross-hybridized only poorly in very low stringency conditions with HIV DNA probes^{19,22}. We have therefore designated this West African AIDS virus as HIV type 2 (HIV-1 referring to the AIDS retrovirus previously identified in Central Africa, North America and Europe). More than 20 isolates have so far been made from patients with AIDS and related conditions, mainly originating from west Africa^{20,21}, but also in some Europeans (L.M., unpublished), and epidemiological studies in progress indicate a seroprevalence of 1-2% in some populations of West Africa (F. Brun-Vézinet, personal communication).

HIV-2 appears to be closely related to the simian immunodeficiency viruses (SIV) a group of cytopathic retroviruses whose prototype, STLV-3_{mac}, was identified in captive rhesus monkeys (*Macaca mulatta*) with an AIDS-like disease²³, and was later found to infect other primate species, either wild or in captivity²⁴⁻²⁶. Genetic comparisons of SIV, HIV-1 and HIV-2 may help to elucidate the phylogeny of these viruses and the origins of the recent AIDS epidemics. As these retroviruses share most of their biological properties, the identification of conserved sequences is important to localize the functional domains of the viral proteins and regulating elements, and design new diagnostic and therapeutic tools. We present here the complete nucleotide sequence of HIV-2, the comparison of its proteins with those of HIV-1, and preliminary studies on the regulation of HIV-2 expression.

Nucleotide sequence and LTR analysis

The sequence presented in Fig. 2 is derived from two λ clones corresponding to integrated proviral DNA from the ROD isolate of HIV-2 (ref. 22), obtained in 1985 from an AIDS patient from Cape Verde Islands (offshore Senegal, refs 19, 20). The genome of HIV-2 is 9,671 nucleotides long (in its RNA form), whereas HIV-1 isolates are about 9,200 nucleotides long. This difference is partly explained by the respective sizes of the long terminal repeats (LTRs, see below).

The genetic organization of HIV-2 (shown in Fig. 1) is analogous to that of HIV-1, that is:

5'LTR-gag-pol-central region-env-orf F-3'LTR.

The 'central region', also identified in the ovine lentivirus visna²⁷, contains five major open reading frames (ORFs), four being clearly related to the ORFs of HIV-1 that encode the Q (or *sor*), R, *tat* and *art* (or *trs*) genes of HIV-1 (refs 15-18, 27-31). The fifth, which we designate ORF X, has no obvious counterpart in HIV-1. Alignments of the nucleotide sequences of HIV-1 and 2 show their distant homology (from ~60% for the more conserved *gag* and *pol* genes, to 30-40% for the other viral genes and LTRs). To allow these alignments to be made many insertions and deletions must be introduced into the sequences. We do not find that these insertions are the small duplications that would be characteristic of the recent divergence of retroviral sequences, as was noted among isolates of HIV-1 (ref. 12).

The limits of the LTRs and of their internal U3, R and U5 elements, determined by sequence analysis and some complementary experiments, are shown in Fig. 2. Classically boundthe retroviral LTRs are short inverted repeats (5' CTG-CAG 3') located after a polypurine tract for the 3'LTR, and before a sequence complementary to the 3' end of a transfer RNA that is used as primer by the reverse transcriptase (here, as in HIV-1 and visna virus, a lysine tRNA, refs 15, 27) for the 5' LTR. The R-U5 junction, corresponding to the 3' end of the polyadenylated viral RNA, was previously localized by sequencing oligo(dT)-primed complementary DNA (cDNA) derived from the HIV- 2_{ROD} genome²². The length of U5 + R, and hence the position of the U3-R junction corresponding to the 5' cap site of the viral RNA were deduced from the size of a HIV-2 cDNA synthesized using the endogenous reverse transcriptase activity and the endogenous tRNA^{lys} primer (see Fig. 3). This

^{*} Present address: Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA † To whom correspondence should be addressed.



Fig. 1 Organization of the HIV-2 and HIV-1 genome (BRU isolate, ref. 15). Vertical bars represent the stop codons in the 3 reading frames. Arrows indicate the initiator AUG codons in viral genes or potential genes. *Tat* 1 and 2, *art* 1 and 2 are the open reading frames containing the coding exons of the *tat* and *art* genes.

'strong-stop cDNA' is 302 + / -1 nucleotides long (181 nucleotides in HIV-1, ref. 15). Thus, the U5 element is 125 bp long, U3 is 556 bp and R 173 bp (respectively 82, 456 and 97 bp in HIV-1). All the elements of the HIV-2 LTRs are larger than in HIV-1, and alignment by computer programs shows large insertions and very distant overall homology for the aligned regions²². However, the three Sp1 binding sites identified in HIV-1 (ref. 32), are also present in HIV-2 from nucleotide 9,419 to 9,448 with 17 out of 29 nucleotides homologous to this region of HIV-1. The core enhancers identified in HIV-1 (ref. 33) are present in HIV-2 from nucleotide 9,389 to 9,416: the first is 50% homologous and the second 100% homologous to that in HIV-1 (Fig. 2).

The analysis of the virus-specific $poly(A)^+$ RNA (not shown) from a cell line infected with and continuously producing HIV-2 revealed a pattern of transcription reminiscent of that observed in HIV-1-infected cells: RNA of over 9 kilobases (kb), corresponding to a full-length transcript, and three types of spliced messenger RNA of 5, 4.5 and 2 kb, also observed in HIV-1 (refs 18, 34).

The gag and pol proteins and HIV phylogeny

The gag precursor of HIV-2 has a calculated relative molecular mass of 57,100 (M_r 57.1K), consistent with the p55 antigen²⁰ seen by immunoprecipitation with patient sera, and is probably processed, by analogy with HIV-1, into the proteins designated p16, p26 and p12 (refs 19, 20). By analogy with the p18^{gag} of HIV-1, p16 would be at the amino terminus of gag and precede p26, whose amino terminus has been sequenced (H. Marquardt, personal communication) and starts with the proline residue at position 951. The carboxy-terminal part of the gag precursor encodes a p12 that contains the cysteine-rich consensus of the retroviral nucleic-acid-binding proteins also found twice in the p13^{gag} of HIV-1 (ref. 15). The HIV-2 *pol* ORF could encode the p64 and p36 antigens of HIV-2 (ref. 20) which by analogy correspond to the p68 and p34 (reverse transcriptase and endonuclease, respectively³⁵) of HIV-1.

The gag and pol proteins of HIV-1 and 2 were expected to share large conserved domains, as these HIV-2 proteins can be precipitated by antibodies in sera from patients infected with HIV-1. However, we found that only 58% and 59.4% of the amino acids of gag and pol respectively are identical to the

corresponding HIV-1 products (Table 1a), whereas the more distant isolates of HIV-1 (Zairian and US) show 90 to 95% amino-acid identity in these proteins (Table 1b and ref. 12). Several insertions and deletions have to be introduced in the alignments (data not shown), whereas they are rare in the comparisons of gag and pol genes between HIV-1 isolates. The gag and pol proteins of HIV-2 are no closer to those of the Zairian isolates than to the prototype HIV-1 (BRU isolate) isolated in 1983 from a French patient⁶ probably infected in the USA. Overall, the difference in gag and pol between HIV-1 and HIV-2 is of the same order as that observed among the group of the human T-cell leukaemia viruses (HTLV-I and II) and bovine leukaemia virus (BLV). However, this latter group displays a higher conservation in the envelope, 70% amino-acid identity between HTLV-I and HTLV-II, versus about 42% between HIV-1 and HIV-2 (see below). Alignments of different retroviral pol proteins (Table 1b) confirm that the HIVs form a subgroup that is more related to the lentiviruses visna and equine infectious anaemia virus (EIAV) than to any other human or animal retrovirus.

Homologous domains in env

The envelope glycoproteins of retroviruses are translated from a subgenomic viral mRNA (here probably the transcript of 4.5 kb). Addition of sugar residues (N-linked glycosylation) gives rise to a high- M_r precursor which is processed by proteolytic cleavage. The length of the leader sequence of the HIV-2 glycoprotein cannot be precisely determined by alignment with that of HIV-1 (experimentally found to be 32 amino acids long³⁶) because of a lack of sequence homology (Fig. 4). But the amino terminus of env contains a relatively hydrophobic stretch in the calculated hydropathy plot (not shown) that is probably the signal peptide. The potential cleavage site between the external envelope glycoprotein (120K) and the transmembrane protein (previously thought to be the 36K antigen¹⁹, and now putatively identified as a 40K antigen²⁰) is found at amino acid 505 (Fig. 4) immediately after the Lys-Glu-Lys-Arg sequence. This cleavage site aligns partly to one (Lys-Ala-Lys-Arg) of the two potential cleavage sites found in HIV-1 (the other being located after the Arg-Glu-Lys-Arg stretch). The calculated M_r of the extracellular glycoprotein (EGP) and of the transmembrane protein (TMP) of HIV-2 would be 57K and 41.7K respectively; the discrepancy

CAAGCAAGTTAAGTGTGTGTGCTGCCATCTCTCCTAGTCGCCGCCTGGTCATTCGGTGTTCAGTGAGTAACAAG	. 3500
	Thr Trp II eProAspTrpAspPheValSerThrProProLeuValArgLeuAlaPheAsnLeuValClyAspProIIeProGlyAlaCluThrPheTyrThrAspClySerCysAsnArg GACATOGATCCCAGACTOGGACTTCGTGTGTGCCCGACGACGGCTGGGGGTTGACGGTTGACGGGATCCGTATACCAGGGGACGGGAGGGCGTGGAGGGGGGGG
300 PBS	3700 GinSerLysGluGiyLysAlaGiyTyrYaiThrAspArgGiyLysAspLysValLysLsuGluGinThrThrAsnGinGinAlaGiuLeuGiuAlaPheAlaMetAlaLeuThrAsp Generative of the contraction of the
Gan	
$GGT_{AGA} = GGT_{AGA} + GGT_$	SerGlyProLysValAsn11e11eValAspSerGlnTyrValMetGlyI1eSerAlaSerGlnProThrGluSerGluSerLysI1eValAsnGlnI1eI1eGluGluMetI1eLysLys CTCGGGCTCCAAAAGTTAATATTATAGTAGACTCACAGTATGTAATGGGGATCAGTGCAAGCCAAGCGAACCCAACAGAGTCAGAAAGTGAAAATAGTGAACCACATGATAAAAAA
ValTrpAlaAlaAsnLysLeuAspArgPheGlyLeuAlsGluSerLeuLeuGluSerLysGluGlyCysGlnLys	
GTGTGGGGAGCGAATAAATTGGACAGATTGGGATTAGCAGAGGGCTGTTGGAGTCAAAAGAGGGGTTGTGAAAA 700	GluAlaIleTyrVelAlaTrpVelProAlaHisLysGlyIleGlyGlyAsnGlnGluVelAspHisLeuVelSerGlnGlyIleArgGlnVelLeuPheLeuGluLyeIleGluProAle GQAAGCAATCTATGTTGCATGGGTCCCAGCCCAGAAAGCATGGGGGAAACCAGGAGGAGGGGAGCCCGGGGGGGG
LeuLysSerLeuPheAsnThrValCysValIleTrpCysIleHisAlaGluGluLysValLysAspThrGluGly	
T TAAAA AGTCTTTT TA ATACTCTCTCCCCT CATTTCGTGCATACACCCCAGA AGGAA AGT GAA AGATACTCA AG 800	GINGIUGIUNISGIUCYSIYTHISSEASINVILLYGGIUGEGGA TCAGGAAGAACAIGAAAAATATCATAGCAATGTAAAAGAACTGTCTCATAAATTTGGAATACCCAATTTAGTGGCAAGGCAAATAGTAAACTCATGTGGCCCAATGTCAACGGGGA 4100
AlaGluLysMetProSerThrSerArgProThrAlaProSerSerGluLysGlyGlyAsnTyrProValGlnHis	4100 Allo Sector (Sector) - Allo Sector) - Allo Sector (Sector) - Allo Sector (Sector) - Al
	A IGHTERISON ON THE ASSIGNED LEDGING THE POTTER CAS AND CASE OF SUBJECT AND A GAL CATTATAGTAG CAST ACAT CATTA CAS AND A AND A CASA A
uAsnAlaTroValLvsLeuValGluGluLvaLvaPheGlvAlaGluValValProGlvPheGlnAlaLeuSerGlu	
A ATGCCTGGG TAAA AT TAG TAGAG GAAAAAA AGTTCGGGG CAGA AG TAGTGC CAG GATTT CAGG CACTCT CAG	1] ePr oCl nGl uSerGl y Ar gGl nThr Al ale uPh ele ule ules Le uA la Ser Ar gTrpPr o II eThr Hi ale uHi aThr As pA an Gl y Al a A an PheThr SerGl nGl uVa lles Mat C AT CCAC AGGAATCAGGA AGACA AACAGCACTCTTCCTATTGA AACTGOCA AGTAGGTGGCCGA ATAACACACTTGGATACAGATAATGGTGCCAACTTCACTTGAGAGGGGGAGGGGAAGAT
As pHisGlnAlaAlaMetGlnIleIleArgGluIleIleAsnGluGluAlaAlaGluTrpAspValGlnHisPro	
CGACCATCAAGCAGCCATGCAGATAATCAGGGAGATTATCAATGAGGAAGCAGCAGCAGGAATGGGATGTGCAACATC 1200	Valaiar profileo iy fieo iy fieo iy caberneo iy valro iy ranno on bercino iy valvalu uu lanena kana shi bu eubyea shu hileber ragi ler gu lugin GG tag catgor ga tago ta taga aca arcort taga accort aca arco cacagage ca ab gar taga agoa and an ta ac carc ta aa at a Ac taga aca A con
/SerAsplieAlaGlyThrThrSerThrValGluGluGlnIleGlnTrpMetPheArgProGlnAsnProVelPro	A la ten Fin Fin a Cumberlia Valla u Martala II a Mia Cua Martan Dhai ya ten den CivCiv Jia Civ ten Marthe Pro Sav Civ ten Carlau II a ten Marti a The The Civ Civ
	GCCAAATACAATACAATACAATACTACTAATGCCAATTCATTC
TOTOTOTAGGATGTACAACCCGACCAACATCCTAGACATGAGACACCGGGCGAACGGCCCTTCCAAACCTATC	GlulleGlnPheLeuGlnAlalysAsnSerLysLeuLysAspPheArgValTyrPheArgCluGlyArgAspGlnLeuTrpLysGlyProGlyGluLeuLeuTrpLysGlyGluGlyAla
	AGA GA TA GA ATTOCTO CA AGO CAAAA ATTOAAA AT TAAA AGATTTTOGGOTO TATTT CAGAGA AGG CAGAGAT CAGTTGTG GAA AG GAOCTGGG GA AC TACTGTG GAA AG GAGA AG GAOC
aValLysAsnTrpMetThrGlnThrLeuLeuValGlnAsnAlaAsnProAspCysLysLeuValLeuLysGlyLeu	. 4700
NGT GA AGA ATTG GAT GACC CAA ACACTGC TAG TACAAA ATGC CA ACC CAGACTG TAA AT TAGTGC TAAA AG GAC 1500	ValLeuValLysValGlyThrAsplieLysIIelleProArgArgLysAlaLysIIelleArgAspTyrGlyGlyArgGInGluKetAapSerGlySerHisLeuGluGlyAiAArgGlu Q MetGluGluAspLysArgTrpIleValValProThrTrpArgValProGlyArg
ValGlyGlyProGlyGlnLysAlaArgLeuMetAlaGluAlaLeuLysGluValIleGlyProAlaProIlePro	AGTEC TAGT CA AGG TAG GA ACAGA CA TAAAA ATA ATACA AG GAA AG GAA AGC CA AGAT CAT CAGAGAC TATG GAG GA AGACA AGAGATG GATAGTGGTTGC CACGTG GAGGGTGC CAGG GA
GGTAGGTGGGCCAGGCCAGAAAGCTAGATTAATGGCAGAGGCCCTGAAAGAGGTCATAGGACCTGCCCCTATCC	
	aspujoune calay Merciul verenki se zi zuvali vervri za Tvri večani zučiul veval Cvervrval Prokiski si veval Givrral arrotrotrotrove Serároval Tie
IG GAA AG GA AGGG CACTCGG CA AGACA ATGCC GAG CACCTAGA AGGCAGGGCTGCTG GA AGTGTGG TA AGC CAG	G GATE GACAA ATEGE CA TAGECTTET CA AG TATE TAAA ATACAAA AGATE TAGAAA AGETETECT TATETTECEE CACE ATA AGETEGE GATEGE CATEGE GACTTE CAC CACE ATA
The Civer of the C	PheProLeu(vsG)vAsnSerHisteuG)ul]eG nA]aTvrTrpAsnLeuThrProG uLysG)vTrpLeuSerSerTvrSerValAralleThrTrpTvrThrGuLysPheTrpThr
vLeuGlyProTroGlyLyslysProArgAspPheProVs]AlaGlyValProGloGlyLeuThrProThrAlsPro	I AT TECCATTAAAAGGA AACAGTCATCTAGAGATACAGGCATATTGGAACTTAACACCAGAAAAAGGATGGCTETEETTATTCAGTAAGAATAACTTGGTACACAGAAAAGTTCTGGA
ACTEGEECCTTEGE GAA AGA AGCCCCCGCA ACTTCCCCCGTEGCC CA AGTTCCG CAGGEGECT GACAC CA A CAG CAC	
	AspValThrProAspCysAlaAspValLeuIleHisSerThrTyrPheProCysPheThrAlaGlyGluValArgArgAlaIleArgGlyGluLysLeuLeuSerCysCysAsnTyrPro
Ar gGl uLysThrGl uAr gAlaGl uAr gGl uThr II eGl nGl ySerAs pAr gGl yLe uThrAlaPr oAr gAlaGl y	CAGATGT TACCCCAGACTGTG CAGATGTCC TA ATACACTTATTTCCCTTGCTT TACAG CAGGT GA AG TA AGA AGAGC CAT CAGAGGG GAAA AGTTATTGTCCTGCTGCAATTATC
nClyLysArgGlnArgCluGlnArgCluArgProTyrLysCluYalThrGluAspLeuLeuHisLeuGluGlnGly	X Met Thr As oProAreGluThr VelProProGlyAsoSerGlyGluGluThr I eGly
2000	ArgAlaHisArgAlaCinValProSerLeuCinPheLeuAlaleuValValValGinGinAsnAspArgProGinArgAspSerThrThrArgLysGinArgArgArgArgArgArgArg
GinPheSerieuTrpLysArgProValValThrAlaTyrIleGluGlyGlnProValGluValLeuLeuAspThr	CCC GÁGCT CATAGAGCO CAGG TACCGT CACTT CA ATTTCTGGCCT TAGTGG TAGTG CA ACAAA AT GA CAGÁGCCC CAGAGÁGA CAGTAC CAC CAG GAA A CAGCGGC GA AGÁGAC TÁTCG GA
uAsnSerLeuPheGlyLysAspGln.	. 5300
CA ATTCTCTCTTTG GAAA AGAC CAG TAGT CACAG CA TACATT GAGGGT CAGC CAG TAGA AGTCTTGT TAGACA C	GluAlaPheAlaTrpLeuAsnArgThrValGluAlaileAsnArgGluAlaValAsnHisLeuPrOArgGluLeuIlePheGlnValTrpGlnArgSetTrpArgTyTrpHisAspGlu GluAlaPheAlaTrpLeuAsnArgThrValGluAlaileAsnArgGluAlaValAsnHisLeuPrOArgGluLeuIlePheGlnValTrpGlnArgSetTrpArgTyTrpHisAspGlu
	Gy beung geun aby so that poet negoet negoet negoet aby so that goet set of to set for the reading and the reading
I yr Ser Prolysil evalgiygiyil egiygiyrh eil eAsnThrlysgiuTyrlysAsnValgiuII egiuVal	
	GinGiyMetSerGluSerTyrThrLysTyrArgTyrLeuCysIleIieGlnLysAlaValTyrMetHisValArgLysGlyCysThrCysLeuGlyArgGlyHisGlyProGlyGlyTrp
IleAsnIlePheGlyArgAsnIleLeuThrAlaLeuGlyMetSerLeuAsnLeuProValAlaLysValGluPro	CAAGGGATGTCAGAAAGTTACACAAAGTATAGATATTTGTCGATAATACAGAAAGCAGTGTACATGCATG
AL CANUALI LI LIGUA GANATATTUT GAUAGUUT TAGGUATGT CATTAA ATU TAU CAGTUGU CAA AG TAGAGUU	ArgProGlyProProProProProProClyLeuVal
CinTrpProLeuThrLysGluLysIleGluAlaLeuLysGluIleCysGluLysMetGluLysMetGluCivCivCivCiv	🗧 🦰 Me talaGlualaProThrGluLeuProProValaspGlyThrProLeuArgGluProGlyAspGluTrpIleIleGluIle
CAATGGCCCTTAACAAAAGAAAAAATAGAAGCACTAAAAGAAATCTGTCAAAAAATGGAAAAAGAAGGCCAGCT	AGACCAGGGCCTCCTCCTCCCCCCTCCAGGICIGGTCTAATGGCTGAAGCACCCACGAGGGCCCGGGTGGATGGGCCCGGCTGGAGGGGGGGCCGGGGGGGG
IleLysLysLysAspLysAspLysArpArgMetLeuIleAspPheArgGluLeuAspLysValThrGlnAspPhe	101 🖊 metotuumin toetuu yhteen kan kan kan kan kan kan kan kan kan ka
AT CA AGAAAA AG GACAAAA ACAA ATG GAG GATGC TA ATAGATTT CAGAGA AC TAA ACA AGG TA ACT CA AGATTT	T CTTGAGAGAAATAAAAGAAGAAGATTTAAAGGATTTTGACCCTCGCTTGCTAATTGCTCTTGCGAAATATATCTATACTAGACATGGAGACACCCTTGAAGGGCGCCAGAGAGGCCCATTA
AAGAGAAGAATTACTGTACTAGATGTAGGGGATGCTTACTTTTCCATACCACTACATGAGGACTTTAGACCATA	LysSerCysAsnGluProPheSerArgThrSerGluGlnAspValAlaThrGlnGluLeuAlaArgGlnGlyGluGluIleLeuSerGlnLeuTyrArgProLeuGluThrCysAsnAan
. 2700	Val LeuGlnArgAla LeuPheThrHisPheArgAlaGlyCysGlyHisSerArgIleGlyGlnThrArgGlyGlyAsnProLeuSerAlaIleProThrProArgAsnHetGln 🖷
ArgTyrIleTyrLysValLeuProGlnGlyTrpLysGlySerProAlallePheGlnHisThrMetArgGlnVal	A AGTECTGEAACEAGECECTTEACGAETTEACGAETTEACGAEGAAGGATGTGGECACTEAAGAATTGGECAAGAAACEGGGAGGAAATCETCTETCAGGAEGCAECECETAGAAAACATGGAATAAC
A GATA CATA TATA AGT CTTCC CA CAGG GATG GA AGG GAT CAC CAG CA ATTTTT CA A CA CA CA AT GAGA CAGGT	2700 Sarfur Tur fur fur fur fur fur fur fur fur fur f
	art Hettanglukrahaspillelike versionen and hetter hettanglukrahaspillelike versionen and hettan
TACATGGATGATATCTTAATAGCTAGTGACAGGACAGATTTAGAACATGATAGGCTAGTCCTCCTCCACCA	ICALGCTATIGTAAGCGATGCTGCTACCALIGTCAGATGTGTTTTCTAAACAAGGGGCTCGGGATATGTTATGAACGAAAGGCCAGAGGAAGAAGAACTAAGAAAGTAGA
3000	
LysAspProProTyrHisTrpMetGlyTyrGluLeuTrpProThrLysTrpLysLeuGlnLysIleGlnLeuPro	ProSerProThrProAspLys
AA AGACCETE CATAC CACTE GATEGEE TAT GA AC TATEGEE CA AC TAA ATE GA ACTTE CAGAAA ATACAGTTECE	Argle ole uti süininr
	$\blacksquare H = exect a construction of the sector is the executive sector is the sector of the sector is th$
VAILEUASDITEAIAGIDLEUTYPProGlylleLysThrLysHisLeuCysArgLeuIleArgGlyLysMet	
3200	LysAsnAlaThr II eProLeuPheCysAlaThrArgAsnArgAspThrTrpGlyThr II eGinCysLeuProAspAsnAspAspTyrGinGiuIi eThrLeuAanValThrGluAlaPhe
LeuGluGluAsnArglielleLeuSerGinGluGlnGluGlyHisTyrTyrGinGluGluLysGluLouCluAta	G CA AAAATGCAACCATTCCCCTCTTTTGTGCAACCAGAAATAGGGATACTTGGGGAACCATACAGTGCTTGCCTGACAATGATGATTATCAGGAAATAACTTTGAATGATGATGATGACGAGGCTT
C TA GA A GAAA A CA GA AT TATCC TA AGC CAG GA ACA AGAGG GA CAC TAT TAC CA AGA AGAAAA A GAGC TAGA AGC	
	AspAlaTpAsnAsnThrValThrGluGlnAlalleGluAspValTrpHisLeuPheGluThrSerlleLysProCysValLysLeuThrProleuCysValAlmeLysCysSerSer I GATECAICDAINA TACAGTAACACAACAACACACATACAACAGTCTCTCCAICTATTCCGGGUTCAITAILGTCTTCCTICTICTICTICTICTICTICTICTICTIC
GINGIUGIULys II eLeuLys ValGlyLysTyrAlaLys ValLysAan ThrHis ThrAanGlyII eAr gLeuLeu	ALOO

AlaGInValValGInLysIieGlyLysGluAlaLeuValIIeTrpGlyArgIIeProLysPheHisLeuProValGluArgGluIleTrpGluGlnTrpTrpAspAsnTyrTrpGlnVal AG CACAGG TAGTT CA GAAA A TAG GAAA AGA AG CAC TAGT CATTTOGG GAC GA ATACCAAAATTT CACCTACCAG TAGAGAGAGAA ATCTGG GAG CAGTGGTGG GATAAC TACTGG CA AGT

1.15 100 GCCCCACGCTTGCTTGCTTAAAAACCTCTTAATAAAGCTGCCAGTT ACCCTGGTCTGTTAGGACCCTTCTTGCTTTGGGAAACCGAGGCAGGA GGCAGTAAGGGCGGCAGGAACAAACCACGACGGAGTGCTCCTAGAAA 400 . . . ACAC CAAAA ACTG TAGCC GAA AGGGCTTGC TATCC TACCTT TAGACA 500 ArgleuArgProGlyGlyLysLysLysTyrArgLeuLysHisIle T CAGGTTACGGCCCGGCGGAAAGAAAAAGTACAGGCTAAAACATATT IleLeuThrValLeuAspProMetValProThrGlySerGluAsn AA ATTCT TACAGTTT TAGATC CA ATGG TACC GACAGGTT CAGAAA AT AlaLysGinIleValArgArgHisLeuValAlaGluThrGlyThr GAGCAAAACAAATAGTGCGGAGACATCTAGTGGCAGAAACAGGAACT ValGlyGlyAsnTyrThrHisIleProLeuSerProArgThrLeu ATGTAGGGGGAACTACACCCATATACCGGTGAGTGCCCCGAACCCTA 1000 GlyCysThrProTyrAsplleAsnGlnMetLeuAsnCysValGly AAGGCTGCACGCCCTATGATATCAACCAAATGCTTAATTGTGTGGGG 1100 IleProGlyProLeuProAlaGlyGlnLeuArgGluProArgGly CA A TAC CAGGCCCCT TAC CAGCGGGG CAGCT TAGAGAGC CA AGGG GA ValGlyAsnIleTyrArgArgTrpIleGlnIleGlyLeuGlnLys CAGTAGGAAACATCTATAGAAGATGGATCCAGATAGGATTGCAGAAG AspArgPheTyrLysSerLeuArgAlaGluGlnThrAspProAla GlyMetAsnProThrLeuGluGluMetLeuThrAlaCysGlnGl I AGGGATGAACCCTACCTTAGAAGAGATGCTGACCGCCTGTCAGGGG 1600

PheAlaAlaAlaGlnGlnArgLysAlaPheLysCysTrpAsnCys CATTCGCAGCAGCCCACCAGAGAAAGGCATTTAAATGCTGGAACTGT 1700

Pol > ThrGlyArgPhePheArg? HisIleMetThrAsnCysProAspArgGlnAlaGlyPheLeuGl GACACAT CAT GACAA ACTGCC CAGATAGACAGG CAGGTTTTTTAGG

ProSerGlySerSerSerGlySerThrGlyGluIleTyrAlaAla ProValAspProAlaValAspLeuLeuGluLysTyrMetGlnGl CCCCAGTGGATCCAGCAGTGGATCTACTGGAGAAATATATGCAGCA

GlyAspThrIleGlnGlyAlaThrAsnArgGlyLeuAlaAlaPro GluThrProfyrArgGluProProThrGluAspLeuLeuHisLe GGGAGACACCATACAGGGAGCCACCA ACAGAGGACTTGCTGCACCT

GlyAlaAspAspSerIleValAlaGlyIleGluLeuGlyAsnAsn

AGGGGCTGACGACTCAATAGTAGCACGAATAGAGTTAGGGAACAAT 2200 2200 LeuAsnLysLysValArgAlaThrIleMetThrGlyAspThrProl ICTAAATAAAAAGGTACGGGCCACCATAATGACAGGGGACACCCCA

2300 IleLyslleMetLeuLysProGlyLysAspGlyProLysLeuArg AATAAAAATAATGCTAAAGCCAGGGAAAGATGGACCAAAACTGAGA

GluGluAlaProProThrAsnProFyrAsnThrProThrPheAla AGAGGA AG CACCTC CA ACTA ATCCTTA TA A TACCCCC CACATTTG CA.

ThrGluIleGlnLeuGlyIleProHisProAlaGlyLeuAlaLys CACAGAAATICAGITAGGAATICCACACCCAGCAGGGTTGGCCAAGA

ThrAlaPheThrLeuProSerValAsnAsnAlaGluProGlyLys TACTG CATT TACTC TAC CAT CAGT CA ACA ATG CAGA AC CAG GAAAA 2800 LeuGluProPheArgLysAlaAsnLysAspValIleIleIleCln

LeuleusanGiyleuGiyPheserThrPromspGiulysheGin LeuleusanGiyleuGiyPheSerThrPromspGiulysPheGin AcitcTanatGCcTaGaatatcc

GInLysGluIleTrpThrValAsnAspIleGinLysLeuValGly CCAAAAAGAAATATGGACAGTCAATGACATCCAGAAGCTAGTGGGT

ThrLeuThrGluGluValGlnTrpThrGluLeuAlaGluAlaGlu GACACT CACAGA AGA AGTACAGTG GACAGA AT TAG CAGA AG CAGAGG

Thrvaiginlysaspoinciussoginciussogincitaspoincity lysileniscincius vaigus vaigu . . 3400

© 1987 Nature Publishing Group

Thr GluSerSerThr GlyAsnAsnThr Thr SerLysSerThr SerThr Thr Thr Thr Thr Thr Thr Thr Thr GluGluGluGluGluGluGluGleGradAsnThr Thr SerLysSerThr SerThr Thr Thr Thr Thr Thr Thr Thr Thr Thr	•
0000 Ser GlyLeuglyGluGluGluThrileAsnCysGlnPheAsnMetThrGlyLeugluArgAspLysLysLysGlnTyrAsnGluThrTrpTyrSerLysAspValValCysGluThrAs SCTCAGGAITGGGAGAGGAGAAAACGATGATCAATGGCAGTTCAATATGACAGGATTAGAAAGGATAAGAAAAAAGAGTATAATGAAACATGGTGCTGTGGTGGTTGTGAGACAA	n
Asn Ser Thr AsnG In Thr GInCys Tyr NetAsnHis Cys Asn Thr Ser Val 11 e Thr GI uSer Cys AspLys His Tyr Tr pAspAlalle Arg Phar Egyr Cys AlaPro ProGly Ty ATA A TAG CACA AT CAGACCCAGTG TACATGA AC CATTG GA AC CATTGGA AC GAGT CATGACAAG CAC TATGGGATGCTATA AGGTT TAGATACTGTG CAC CACCGGGGT AGO	r
Al aleule uAr gCysAsnAspThrAsnTyrSerClyPheAl aProAsnCysSerLysValValAl aSerThrCysThrArgHetHetCluThrGinThrSerThrTrpPheGlyPheAa A TGCCCTAITAAGATGTAATGATACCAATTATTCAGGCTTTGGACCCAACTGTTCTAAGTAGGTAG	n
GlyThrargalaGluAsnArgThrTyrIleTyrTrpHisGlyArgAsnArgThrIleIleSerLeuAsnLysTyrTyrAsnLeuSerLeuHisCysLysArgProGlyAsnLysTh AIGGCACTACAGCAGAGAATACAACATATATCTATTGCAGCGCAGGGCAATAGGACTATCATCACCTACAGCTACAACAATATTATAATCTCAGTTTGCATTGTAAGAGGCCAGGGAATAAGA	IT
VallysGinlieMetLeuMetSerGiyHisValPheHisSerHisTytGinProIleAsnLysArgProArgGinAleTrpCysTrpPheLysGiyLysTrpLysAspAlaMetGinGi CAGTGAAACAAATAATGCTTATGTGAGAGCATGTGTTCACTCCCACTAGCAGCGGCGGTCAATAAAAGAGCCCGGAGAAAGGAGGGGGTGGTGGTGGTGGTGGTGGTGGAAA 7100	u
VailysGluThrLeuAlalysHisProArgTyrArgGlyThrAsnAspThrArgAsnIleSerPheAlaAtaProGlyLysGlySerAspProGluVaiAiaTyrHetTrpThrAsnCy AGGTGAAGGAACCCITGGAAAACATCCCAGGTATAGAGGAACCAATGACACAGGGATATTAGCTTTGCAGCGCCAGGAAAAGGCTCAGACCCAGAACATGCGACCAGGA 7300	
ArgGlyCluPheLeuTyrCysAsnMetThrTrpPheLeuAsnTrpIleGluAsnTysThrHisArgAsnTyrAlaProCysHisIleLysGlnIleLleannThrTrpHisLysValOl JCAGAGGAGAGATTICICTACTGGAACATGACTTGGTTCGTCATTGGATAGAGAATAAGGACACCGCGCAATTATGCACCGTGCCGTAATAAAGGAAATAATTAACACATGGCATAAGGTAG	у
Argasn Val Tyr Leu Pro ProargGluGlyGlu Leu SerCysaan SerThr Val Thr Ser I I ei I ea I ansn I I ea b ptrpGlnaan an ann an ann an thrasn I I ethr Phe Ser Ai eGI GGA GAAATG TATAI I TGC TGC CAGC GA AGGG GAGCT GTCCT GGAACT GA AGGT ATAGC TA AGATT GACTAG GAAAACAATAAT CAGAGAA ACATT AGCT TAGT GGAG 7500	u
ValAlaGI ULEUJYTAT BLEUGI ULEUGI YAS pTYTLYSLEUVA IGI UII ETHTPTOII EGI YPHEAL APT OTHTLYSGI ULYSAT BTYTSET SETALAHI SGI YAT BHI STHTAT BGI AGGTGGGAGAACTAIACAGAT TGGAGTIGGAGATATATAAT IGGTAGAAATAACACCAATTGGCTTCGCACCTACAAAAGAAAAAGATACTCCTCTGCTCACGGGGGACATACAAGAG 7600	у
ValPheVelLeuGlyPheLeuGlyPheLeuAlaThrAlaGlySerAlaMetGlyAlaAlaSerLeuThrValSerAlaGlnSerArgThrLeuLeuAlaGlyIleValGlnGlnGlnGlnGl GTGTGTTGCGTTGGGTTGTGGGTTTTGCGACAGGGGGGGG	n
Ginleuleuas pValValLysArgGinGinGiuleuleuArgleuThrValTrpGiyThrLysAsnleuGinAlaArgValThrAlaIIeGiuLysTyrleuGinAasgInAlaArgLa AACAGCTGTTGACGTGGTCAAGAGCAAGAACAAGAACTGTTGCGACTGACCGTCTGGGGAAGAACCATGGCAAGAGTGACTGAC	u
AsnSerTrpGlyCysAlaPheArgGlnValCysHisThrThrValProTrpValAsnAspSerLeuAlaProAspTrpAspAsnMetThrTrpGinGluTrpGluLysGlnValArgTy TAAATTCATGGGGATGTGCGTTTAGGGAAGGCTGGCGAGGATGGGGAAGAGCAGGGGTGAGTGGGGAAAGGCGGGGGAGGGGGGGG	r
Leugiualaasniieserlysserleugiuginalaginiieginginginginginginginginginginginginging	r
IrpValLysTyr11eC1mTyrG1yValLeuI1ei1eValA1aVa1I1eA1aLeuArg11eVa1I1eTyrVa1Va1G1mMetLeuSerArgLeuArgLysG1yTyrArgProVa1PheSe CCTCGGTCAACTATATTCAATATGGAGTGCTTATAATAGTACCGCTAATAGCTTTAAGAATAGTGATATATGTAGTAGAAATGTTAAGTAGGCGCTATAGGCGCTATAGGCCGCTTTTCT 8200	r
tat SerlleSerlteArgThrClyAspSerGlnProThrLysLysClnLysLysThrValGluAlaThrValGluThrAspThrClyProGlyArg art AsnTyTroGluGlyProGlyThrAlsSerClnArgArgAsnArgArgArgArgArgArgArgArgArgArgArgArgArgA) e
8600 SerileTyrThrPheProAspProProAlaAspSerProLeuAspGinThrIleGinHisLeuGInGlyLeuThrIleGinGluLeuProAspProProThrHisLeuProGluSerGin AlaTyrIleHisPheLeuIleArgGinLeuIleArgLeuLeuThrArgLeuTyrSerIleCysArgAspLeuLeuSerArgSerPheLeuThrLeuGinLeuIlETyrGInAsnLeuAr SAGCATATATACATITCCTGATCCGCCACTGATTCGCCTGTTGACGAGACTATACAGGATCTGCAGGGATCTACTATCGAGGAGCTTCCTGACCGCCACTCATCTACCAGAATCTCA	B
ArgleuAlaGiuThre	
AspTrpLeuArgLeuArgThrAlaPheLeuCInTyrClyCysCluTrpIleCInGluAlaPheCInAlaAlaAtgAlaThrArgCluThrLeuAlaClyAlaCysArgClyLeuTr GAGACTGCCIGAGACTTAGAACAGCCTTCTTGCGATATGGGTGCGAGTGGATCCAAGAAGCATTCCAGGCGGCGCGCGGCGGGGGGGCGTCTGGGGGGGG	P
GlyGlyTyrTrpAsnGluSerGlyGlyGluTyrSerArgPheGlnGluGlySerAspArgGluGlnLysSerProSerCyGluGlyArgGlnTyrGlnGlnGlyAspPheHetAsnThr ArgYalte.GluArglleGlyArgGlyTlete.AlaYalProArgArglteArgGlnGlyAlaGluItalate.Leu GAGGGTATTGGArGAATGGARGGAATGAATGGCGTTGACGAGGGGGGGGGGGG	
ProTrolysasoProAlataGiutaroEului vatorolauTustaroLocatata Utata	
CCATGCAAGGACCAGCAGCAAAGGGGGAAAAATATTGTAAGGGCAACAAAATATGGATGATGATGATGATGATGATGACGACGAAGGTTGTGTCACGCGAAGAAGTAGCGACAAAATATTGTAAGGGCAACAAAATATGGATGATGATGATGATGATGACGACGAAGGTTGTGTCACGACGAAAAGTAGCGACGAAGAAGTAGGGGAACAAAATATGGATGATGATGATGATGATGACGACGAAGGGTAGGGGTAGGGGTAGAGGATGATGACGACGAAGAGGTGATGAGGGTAGGGGTAGGGGTGATGAGGGTGATGA	
Argen fone chansas i beunali expersormi slevi i elys intargoly GiybeuGluCi yhe FheyrserGi uargarghi si ys Ii elevasni i etyr levGlutys AgaCaan Kacaan Kacaan 8900 PPT - U3 - 900	
0. GUUL 71 ETTENTAND FFDUTAS AT YFTHETH GCI PF COLYVALAR gTYPF OME THE PH ECLYTFPLEUT FPLYALEUVALPF OVALAA PVALPF CGLAGGLAGGGGGAGGAG GAGGAGGGGATAAT TGCGAGAT TGCGAGAAC TACAC TCATGGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	
Anto rount in substant strok laGinThrSerLysPheAspAspProhisGlyGluThrLeuValTrpGluPheAspProLeuLeuAlaTyrSerTyrGluAlaPheIleArg ACTGAGACTGACHCCCTTGCTGCCTTATAGTACATCCAGCACAAACAAGCAAG	
TyrProGluGluPheGlyHisLys5erGlyLeuProGluGluGluTrpLysAlaArgLeuLysAlaArgGly11eroPheSer TACCCACAGGAATTTCGGCACAAGTCAGGCCTGCCAGAGGAAGAGTGGAAGGCGGAGACTGAAAGCAAGAGGAATACCATTTAGTTAAGACAGGAACAGGTATACTTCGTCAGGGGAGGGA	
ACTAGE TAREAGAAAGAGET GAGGET GGAGGAGET GGAGGAGET GGAGGAGET GGAGGAGET GGGGGAGET GGGGGAGET GGGGGGGGGG	
AGCTIGCATIGTACITCGGTCGCTCTGCGGAGGGCCGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCAGTAGGGCGGGTGGCGGGGTGTCCCTGGCTAGACTCTCACCAGCACTTGG 9500 9600	
UUGIGEIGGEGGCCCCACGCTIGCTIGCTIGCTAAAAACCTCCTTAAAAACCTCCTTAAAAAGCA	

6

1987 Nature Publishing Group

Fig. 2 Complete nucleotide sequence of the HIV-2_{ROD} provirus genome. The sequence of the 9,671 nucleotide coding strand from the cap site to the polyadenylation site of the RNA is shown together with the deduced amino-acid sequence of the viral proteins. The 3-bp inverted repeats bounding the LTR are underlined with arrows. The primer binding site (PBS) complementary to the 3' end of tRNA^{1ys}, and polypurine tract (PPT) are underlined. The repeat of the PPT in the *pol* gene is also underlined. The limits of the U3, R and U5 elements of the LTR are indicated by arrows. The promoter (TATAAA), the three potential Sp1 factor binding sites (indicated by dotted lines) and the two core enhancer sequences (E, indicated by dashed lines) of U3 are shown; in R the polyadenylation signal AATAAA is underlined. The viral genes and potential genes are translated from the first AUG of the ORF. The end of translation is indicated by a filled circle. The amino terminus of the major core protein, the 26^{aog} was determined by a filled circle. The amino terminus of the major solution 8,307 indicate the probable splice donor and acceptor sites of the intron separating the two coding exons of the *tat* and *art* genes.

Methods. The sequence was determined by the M13 shotgun cloning and dideoxynucleotide chain terminator method^{57,58}, as described²⁷ starting from the inserts of the λ phages ROD 27 and ROD 35 containing integrated proviral DNA²². ROD 27 corresponds to the 5' part of the genome, to the *Eco*RI site at position 2,658, whereas ROD 35 corresponds to the part of the genome 3' to this site. An in-frame TAG stop codon was found in ROD 35 *env* gene. The sequencing of the corresponding region in an oligo(dT)-primed HIV-2 cDNA (clone E2, described in ref. 22) revealed that it was due to a C-to-T mutation at position 8,304.

665

305-

291

282-

240

214



Fig. 3 HIV-2 strong-stop cDNA corresponding to the length of the R+U5 elements of the HIV-2 LTR. The methods were previously described¹⁵. Briefly, virions were purified by ultracentrifugation and an endogenous cDNA reaction performed with radiolabelled nucleotides after mild disruption of viral envelope with Triton X-100. The tRNA^{1ys} primer, complementary to the PBS site flanking the U5 element at the 5' end of the genome, was then degraded by alkaline hydrolysis and the cDNA run on a denaturing 6% acrylamide-urea gel together with a sequence reaction for accurate estimation of the size of the products.

with the apparent M_r of the EGP is explained by glycosylation (30 sites in HIV-2, about half of which are conserved with respect to HIV-1).

Figure 4 shows an alignment of the envelopes of the two HIVs. The proteins are overall very distantly related (41.7% identity in the entire envelope, 39.4% in the EGP, 44.8% in the TMP) compared to divergent isolates of HIV-1 (about 75-80% identity in the whole envelope, ref. 12). Many large insertions have to be introduced, particularly in alignment of the EGPs where only short, widely separated domains are conserved between HIV-1 and 2. These domains are clustered into the conserved regions of the EGP of HIV-1 (identified by comparison of different isolates¹²⁻¹⁴), and generally coincide with cysteine residues. Among the HIV-1 isolates, all the cysteine residues could be aligned in spite of the generally large genetic variation, especially in gp110. Almost all (22/23) of the cysteine residues of HIV-1 can also be aligned with HIV-2, but the latter contains seven additional cysteine residues, often in the regions representing insertions relative to HIV-1. Thus, the folding of the HIV-2 EGP could be different from that of HIV-1, and some regions, therefore, might be exposed in a different manner.

Other viral proteins

The HIV-1 genome contains several other genes encoding proteins of small M_r (10 to 27K), two of which (*tat* and *art/trs*) have an identified function: the positive regulation of viral expression³⁰⁻³³. No role has yet been identified for the p23 encoded by ORF Q (or *sor*)^{37,38}, nor for the p27 encoded by ORF F (or 3' ORF)³⁹. We also observed in the region between the *pol* and *env* genes of HIV-1 (central region) another potential gene, which we designated R (ref. 12). All these elements are found in HIV-2, but the corresponding proteins are only distantly homologous (see Table 1*a*). In the F protein, most of the difference between HIV-1 and 2 is due to a large insertion in the amino terminus of HIV-2. The second half of the protein, encoded by the U3 element of the LTR, shows better conservation (data not shown).

Based upon sequence homologies with HIV-1, the tat and art genes of HIV-2 are probably organized as split genes transcribed into $\sim 2 \text{ kb}$ mRNA made of three exons^{18,28-31}: the 5' leader, a first coding exon located in the central region and probably ending at a possible splice donor found at position 6,140 (CAAGT, Fig. 2), and a last exon probably starting at the splice acceptor at position 8,307 in HIV-2 (CAGATC). The tat protein of HIV-2 would be longer than that of HIV-1 (130 versus 86 amino acids), having two large insertions in the amino terminus and in the second coding exon (Fig. 4). The main domain of homology of the tat proteins corresponds to a region very rich in cysteine residues whose structure is reminiscent of that of the 'cysteine fingers' of some transcription-regulating elements that interact with nucleic acids, such as the TFIIIA factor⁴⁰. This region is followed by an Arg-Lys-rich stretch that could also interact with DNA or RNA. No significant homology is seen in the second coding exon, which has been shown to be dispen-sable to the function of the protein^{28,29}. The *art*-encoded protein is shorter in HIV-2 than it is in HIV-1 (100 versus 116 amino acids), and most of its length is encoded by the last exon. The most homologous part is located in a stretch of basic residues that may be able to interact with nucleic acids.

Cross-transactivation of HIV-1 and HIV-2

The trans-activator gene (tat) has been shown to be indispensable for the replication and cytopathicity of HIV-1 (ref. 41).

 Table 1
 Ouantification of the homologies among retroviral proteins

				-		0		0		
a	HIV.1 GAG		POL ENV EGP		NV	TMP F		Central	Region tat art	
	HIV.2	57.7 (95.2)	59.4 (96.6)	39.4 (90.6)	44.8 (91.3)	37.7 (79.7)	34.6 (87.4)	52.2 (85.7)	42.8 (64.6)	44.8 (97)
	b		Н	IV-2	F	HIV-1	ŀ	ITLV-I	v	/ISNA
HIV-1		59.1					ND		ND	
			(96.4)							
	LAV-Eli LAV-Mal EIAV		61.6			94		ND		ND
			(96.1) 59 (95.2) 43.8 (92)		((98.7)				
						92		ND	ND	ND
					(98.7)		ND		167
					((01.5)		ND		(00.8)
	VISNA		(92)		,	42.2	ND			(90.0)
	¥151		(8	87)		(94)		ND		
HTLV-I		34.8			33.3				ND	
			(7	(0.5)	(70.3)				
	HTL	V-II	1	ND		ND		62.8		ND
								(99.5)		
	BLV	BLV		ND		ND		49.5		ND
							1	(93.2)		
	RSV		3	5.9		34.5		38.2		ND
			(7	2.3)	(76.2)		(86.4)		
_			(/	4.5)	(,0.2)	-	(00.4)		

The reference protein of each alignment is that listed at the top of the column. Proteins were aligned using the NUCALN program⁶¹ with following parameters: K-tuple 1, window 20, gap penalty 1. Two results are indicated in each case: the amino-acid identity (%) in the aligned domains (that is, excluding the regions of insertion/deletion), and between parentheses the percentage of the length of reference protein that could be aligned. *a*, Homologies between HIV-1 and HIV-2 proteins. For *env*, the calculation was done for the external glycoprotein (EGP, including the signal peptide, whose length is not exactly known in HIV-2), and the transmembrane protein (TMP). *b*, Comparison of the *pol*-encoded proteins of different retroviruses. LAV.Mal and LAV.Eli are Zairian isolates of HIV-1 (ref. 12); EIAV : equine infectious anaemia virus (sequence communicated by Dr S. Aaronson), and visna virus²⁷ are animal lentiviruses; RSV) Rous sarcoma virus⁶⁵. ND, not determined.

ENV	FGP	
HIV.2	M	79
HIV.1	MRVKEKYQHLWRWGWKHGTMLLGILHICSATEKLWYTYYYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFMMKKNDM	100
HIV.2	TEQALEDWHLFETS IKPCVKLTPL CVANKCSSTESSTONNTTSKSTSTTTTT-PT DQE-QEISED TPCARADHCSGLGEEET IN COF NHTGLERDKK KQ	177
HIV.1	VEQMHEDIISUNDOSLKPCVKLTPLOVSLKCTDLCNATNTNSSTNSSSCOMMEXGEIXNCSENISTSIRGKVQKEYAFFYKLDII	187
HIV.2	YNET-WYSKOVVČETŇŇSTŇOTOČYMNHČŇTSVITESČOKHYWDAIRFRYČAPPGYALLRČŇOTŇYSGFAPŇČSKVVASTČTRNHETOTSTWF-GF	271
HIV.1	PIDNDTTSYTLTSCHTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLL-	266
HIV.2	NOT RAERRTYIYWHCRON-RTII-SLINKYYNLSLHCKREON KTVKOIMLMSGHVENSHYOPINKREROANCOEKG-KWK DANO EVKETLAKHER	362
HIV.1	NGS LAEE EV VI RS ANFTDNAKTI I VQL NOS VEINCTRPHNNTRKSIRIQR GPG RAFVTIG KIGNHRQAHCNIS RAKWNATLKQIA SKLRE	356
HIV.2	YRCTNDTRNISFAAPGKGSDPEVAYMMTNCRGEFLYCNMTWFLNWI	449
HIV.1	QFCNNKTIIFKQSS-GGDPELVTHSFNGGGEFFYCNSTQLENSTWFNSTWSTEGSNNTEGSDTITLPCRIKOFINHVQEVCKAMYAPPISQIEGSSN	453
HIV.2	EGPY TMP VISIIANIDWONNOTNITFSA EVAELYRLELGDY KLVEITPIGFA PTKEKRYSSA HGRHTRGVFVLGFLGFLATAGSA HGAASL TYSAQSRT	542
HIV.1	ITCLLLTRDGGHNNNGSE IF RPGGGDHRDNHRSELYKYKYVKIEPLGVA PTKAKERVVQREKRAVGI-GALFLGFLGA AGSTHGARSHTLTVQARQ	548
HIV.2	LLAGIVQQQQQLLDVVKRQQELLRLTVHGTKNLQARVTAIEKYLQDQARLNSHCČAFRQVČHTTVPHVNDSLAPDHDAHTHQEWEKQVRYLEANIS	638
HIV.1	LLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWNNHTWHENDREINNYTSLIH	648
KIV.2	KSLEQAQIQQEK NNYELQKLNSWDIFC NNFDLTSWYKYIQYGVLIIVAVIALRIVIYVQQMLSRLRKGYRPV-FSSPFGYIQQIHIHKDRQQPANEETEE	737
HIV.1	SLIEESQNQQ EKNEQELLELDKWA SLWNWFNITNWLWYIKIFINIYGGLYGLIRIYFA YLSIYN RYRQCYSPLSFQTHLPTPR CPD RPECIEE	740
HIV.2	DGGSNGGDRYWPWPIAYIHFLIRQLIRLLTRLYSICRDLLSRSFLTLQLIYQNLRDWLRLRTAFLQYGCBWIQEAFQ AAARATRETL- -	824
HIV.1	** ** ** ** ** ** *** *** * *** **** * *	833
HIV.2	AGACRGLWRVLERIGRGILAVPRRIRQGAEIALL 858	
HIV.1	**** * ******* * ** VIEVVQGAGRAIRHIPRRIRQGLERILL 861	
Tat		
HIV.2	METPLKAPESSLKSCNEPFSRTSEQDVATQELARQGEEILSQLYRPLETCNNSCYCKRCCYHCQMCFLNKGLGICYERKGRRRRTPKKTKTHPSPTPD	98
HIV.1	MEPVDPRLEPWKHPGSQPKTAC-TTCYCKKCCFHCQVCFTTKALGISYGRKKRRQRRRPPQGSQTHQVSLSK	71
HIV.2	SISTRIC DSQ PTKKQK KTVE A TVET DT GPG R 130	
HIV.1	* ** QPTSQPRCDP	
Art		
HIV.2	MNER-AD-EECLQRKLELIRLINQTNPYPQOPGTASQRRNRRRWKQRWRQILALADSIYTFPDPPADSPLDQTIQHLQGLTI-	81
HIV.1	MAGRSGDSDEDLLKAVRLIKFLYQSNPPPNPEGTRQARRNRRRWRERQRQIHSISERILSTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQCVGSP	100
HIV.2	QELPDP PTHLPESQRIAET 100	

HIV.1 QILVESPTVL-ESGTKE 116

To examine whether transactivation (a property also shared with the ovine visna lentivirus but not with the related caprine arthritis and encephalitis virus⁴²) exists in HIV-2, we constructed a plasmid, called pHIV2-CAT, containing the bacterial chloramphenicol acetyltransferase (CAT) gene under the control of the U3-R region of HIV-2 (225 bp of U3 and 175 bp of R). To test the transactivation of HIV-2, cells were either infected with HIV-2 or mock-infected, and five days later transfected with either pSVCAT (which contains the CAT gene under control of the SV40 early promoter⁴³) or pHIV2-CAT. At the time of transfection, the cells were not producing virus. Nonetheless, we observed a substantial increase in the amount of CAT expression in extracts of HIV-2-infected versus mock-infected cells that had been transfected with pHIV2-CAT (Fig. 5a). The expression of the SV40 early promoter was not affected by HIV-2 infection.

To determine whether the *tat* gene of HIV-1 could transactivate the LTR of HIV-2 and vice versa, we cotransfected SW480 cells⁴⁴ with subgenomic fragments of HIV-1 or HIV-2 and pHIV2-CAT or a plasmid called pHIV1-CAT, which contains U3-R of HIV-1 (the entire U3 and 70 bp of R) directing transcription of the CAT gene. The plasmid pLET (a gift from Dr S. Wain-Hobson) contains the region of the HIV-1 shown by others to encode the HIV-1 *tat* gene^{28,29}. The plasmid pME214, on the other hand, contains HIV-2 sequences between nucleotides 5,786 and 8,571 (Fig. 2), and in particular contains the open reading frames of HIV-2 that share homology with the *tat* gene of HIV-1. In both of these plasmids transcription is driven

Fig. 4 Alignments of the HIV-1 (BRU isolate, ref. 15) and HIV-2 proteins. Asterisks indicate amino-acid identities. Gaps were introduced to optimize the alignments. In the envelopes, the potential cleavage sites are shown by arrows. EGP, external glycoprotein; TMP, transmembrane protein. \diamond , Potential *N*-glycosylation sites; \bullet , cysteines. The domains of the EGP of HIV-1 that were found to be well-conserved among isolates¹² are underlined. The parts of *tat* and *art* encoded by each of the two exons are separated by an arrow.

by the LTR of the respective virus, and the first AUG of the transcript is the first AUG of the putative *tat* gene. It should be noted that both these plasmids also contain the coding potential for the *art* gene.

Although the SV40 early promoter was not affected by either the HIV-1 *tat* nor the HIV-2 *tat* genes, both HIV-1 and HIV-2 LTRs were substantially activated by the HIV-1 *tat* gene (Fig. 5b). This is perhaps surprising in view of the difference in size of the R region of HIV-1 (where the transactivator responsive region (TAR) resides⁴⁵) and HIV-2. However 35 of the 58 bases present in the first stem-and-loop secondary structure of the TAR region of HIV-1 are conserved, and an analogous stem-and-loop structure with the first 77 bases of R can be drawn for HIV-2 (ref. 33).

The HIV-2 LTR is transactivated over 100-fold by pME214 (Fig. 5b). On the other hand, the HIV-1 LTR is not as well transactivated by this plasmid (\sim 5-20 fold, Fig. 5 and other data not shown). Similar results were obtained after transfection of HeLa and HUT 78 cells (data not shown). These experiments indicate that pME214 encodes a functional *tat* gene. In addition, they indicate that the specificity of the HIV-2 *tat* is somewhat different from that of the HIV-1 *tat*. It will be important to determine whether this observation is isolate-specific.

Origin of human immunodeficiency viruses

We have presented here the complete nucleotide sequence of the retrovirus associated with AIDS in West Africa, HIV-2, and tentatively identified the viral proteins either detected in



668

HIV-2. Fig. 5 Transactivation Chloramphenicol of acetyltransferase (CAT) assays were done as described⁵⁹. The unreacted chloramphenicol is marked 'CAM', and the acetylated products are marked 'AcCAM'. All reactions were 1 h with 10% of the cellular extract made 40 h after transfection. The origin of the promoter linked to the CAT gene is indicated above each lane. SV40 indicates the SV40 early promoter, HIV-2 indicates the partial U3 and the entire R sequences of HIV-2 (ROD isolate), and HIV-1 indicates the entire U3 and 70 bp of R of HIV-1 (BRU isolate). a, HUT 78 cells were either mock-infected (UN, uninfected) or infected (IN) with HIV-2. Five days post-infection, 3×10^6 cells were transfected with 3 µg of plasmid in 0.5 ml of Tris-saline without divalent cations for 45 min at 37 °C with 250 µg ml⁻¹ DEAE-Dextran. **b**, 4×10^5 SW480 cells were cotransfected by the CaCl₂ technique⁶⁰ with 3 µg of promoter-CAT plasmid and 3 µg of the indicated plasmid. Salmon sperm DNA was added such that each transfection was 20 μ g ml⁻¹ DNA. This experiment was repeated three times with similar results.

immunoprecipitations with patients' sera, or homologous to proteins previously identified in HIV-1. The two viruses share a similar genomic organization, indicating a common evolutionary origin, but differ significantly in terms of nucleotide and amino-acid sequence: the more-conserved gag and pol genes respectively display only 56 and 60% nucleotide sequence homology and both less than 60% amino-acid identity. The calculation of the nucleotide sequence homology for the other genes gives even lower values, making HIV-1 and 2 42% homologous overall. This confirms that these two viruses are distinct elements of the HIV family, and cannot be considered as strains of the same virus, according to the recommendations of the international taxonomy committee⁴⁶.

It was previously established that HIV-2 is more related to the simian immunodeficiency viruses (SIV) than it is to HIV-1. The gag, pol and env proteins of SIV and HIV-2 are antigenically cross-reactive, whereas their cross-reactivity to HIV-1 is restricted to some gag and pol antigens. The amino-terminal aminoacid sequence of the major core protein (corresponding to the p25gag of HIV-1 and p26gag of HIV-2) has been determined in one isolate of SIV obtained from macaques with an AIDS-like disease (MnIV, ref. 26). Out of the 23 amino acids sequenced 21 match with the amino terminus of p26gag of HIV-2, whereas 13 (with one deletion) match to the p25gag of HIV-1. Furthermore, whereas HIV-2 can infect, at least transiently, primate species which are evolutionarily more distantly related to humans (at least baboons and macaques), HIV-1 infects only humans and chimpanzees (R. Desrosiers and P. Fultz, personal communications). In fact, it is not possible from current data to know whether SIV can be classified as distinct from HIV-2 or if they only differ as independent isolates of the same virus.

The almost simultaneous emergence of two foci of AIDS in distinct areas of the African continent is unlikely to be due to the recent emergence of two novel human pathogens, for example by simultaneous trans-species infection by animal retrovirus, or by the mutation of pre-existing non-pathogenic human retroviruses. Indeed, HIV-1 and HIV-2 are obviously retroviruses with a common origin, but they are highly divergent, and it is more likely that their time of divergence is earlier than the beginning of the current epidemics. Therefore a common ancestor, with similar properties and pathogenic potential, probably existed a long time ago in a human population, and the emergence of the AIDS epidemics is more likely the result of simultaneous modifications of epidemiological parameters in West and Central Africa, such as uncontrolled urbanization, leading to the infection of larger populations.

A question to be addressed is why the HIVs were only recently detected if they existed for a long period. This may be due to the fact that the pathogenicity of an HIV-type retrovirus cannot be revealed until it has spread to a significant portion of the population. First, in areas of Africa with poor medical facilities, where other infections, such as malaria, represent primary causes of morbidity, isolated cases of AIDS could have been an undetectable clinical event. Then, the incubation time can vary considerably, and it cannot still be ruled out that a large fraction of individuals infected by a HIV will remain healthy carriers. In Kenya, HIV-1 seropositivity was first reported in a high fraction of subjects at risk of AIDS (female prostitutes) who were apparently healthy; later, the virus diffused to a larger part of the population, and cases of AIDS were observed⁴⁷. A similar situation could explain the apparent lack of pathogenicity of the retrovirus designated HTLV-IV, but indistinguishable from HIV-2 and SIV by the antigenicity of its proteins^{19,48,49}. The presence of HTLV-IV was identified only in apparently healthy individuals in West Africa, an area where we have observed several typical AIDS cases caused by HIV-2. It is possible that the apparent non-pathogenicity of HTLV-4 is due to a recent epidemic diffusion of HIV-2/HTLV-IV in the West Africa, where AIDS cases still represent a minor fraction of the infected and seropositive individuals, whereas HIV-1 has diffused in major cities of central Africa or the USA some time before.

Implications for vaccines and diagnostics

The risk that HIV-2-infected blood samples may not be detected by standard screens, currently based on the detection of anti-HIV-1 antibodies, makes it important that a way of diagnosing HIV-2 infection is found. As the envelope, and especially its transmembrane part, represents the primary target of the host antibody response to the HIV infection (see ref. 1), antigens from the envelope of HIV-2 will significantly improve the spectrum of the screening tests, allowing the detection of samples infected by HIV-2, and perhaps by other as yet uncharacterized members of the HIV family.

As it shares most of the structural characteristics and biological properties of HIV-1, but displays significant genetic divergence, HIV-2 is a powerful tool in the study of the molecular biology of this group of retroviruses. Among the crucial biological properties common to both HIVs are tropism for CD4positive cells, and mechanisms of positive regulation of viral expression encoded by viral transactivating factors. We observed that the *tat* of HIV-1 activates the transactivation responsive (TAR) sequences as efficiently in both types of HIV, whereas the *tat* gene of HIV-2 is more efficient on the TAR elements of HIV-2. The *tat* proteins of HIV-1 and 2 have only short homologous sequences, and this will ease the dissection of their function by mutagenesis or using chemically synthesized peptides.

HIV-1 and probably HIV-2 recognize the CD4 surface molecule as a receptor on helper/inducer T lymphocytes and perhaps on other cells expressing the CD4 protein⁵⁰⁻⁵³. In HIV-1, this interaction is mediated by the external envelope glycoprotein (EGP; ref. 52), and an important problem is which of the domain(s) of this protein are involved in that interaction. Indeed, blocking this step of the virus life cycle, either by antibodies or drugs, could be an efficient means for preventing infection or blocking its spread. As the receptor is a constant cellular protein, we can postulate that the binding domain of the envelope is conserved among the CD4-tropic HIVs. The conserved domains of the EGP of HIV-1 and 2 are not numerous, and therefore it becomes possible to demonstrate their possible role in the virus-receptor interaction using a relatively limited set of site-directed mutations. Given the absence of antigenic cross-reactivity of the envelopes of the two HIVs, this CD4binding domain is probably not, or only poorly, immunogenicperhaps because of masking by glycosylation, poor exposure on the virion surface, or mimicking of 'self' antigens. Nevertheless, its presentation to the immune system out of context of the virion, that is, as a peptide, might induce a neutralizing antibody response that is not attained, or attained with only a low efficiency, with the complete native envelope from virions or expression systems⁵⁴⁻⁵⁶.

Conclusion

The comparative analysis of HIV-1 and 2 reveals major genetic differences between retroviruses that share many of their biological properties. They both cause AIDS, are cytopathic in vitro,

Received 4 March; accepted 19 March 1987.

- Weiss, R. A. in *RNA Tumor Viruses* (eds Weiss, R., Teich, N., Varmus, H. & Coffin, J.) Vol. II, 405-485 (Cold Spring Harbor Laboratory, New York, 1985).
- Curran, J. W. et al. Science 229, 1352-1357 (1985). 2
- Quinn, T.-C., Mann, J. M., Curran, J. W. & Piot, P. Science 234, 955-963 (1986).
- Piot, P. et al. Lancet ii, 65-69 (1984).
- Clumeck, N. et al. N. Engl. J. Med. 310, 492-497 (1984)
- Barré-Sinoussi, F. et al. Science 220, 868-870 (1983)
- Donovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. Science 224, 497-500 (1984).
 Levy, J. A. *et al. Science* 225, 840-842 (1984).
 Brun-Vézinet, F. *et al. Science* 226, 453-456 (1984). 2
- Ellrodt, A. et al. Lancet i, 1383-1385 (1984) 10.
- Bailey, A. C. et al. Lancet i, 359-361 (1985).
 Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. Cell 46, 63-74 (1986).
 Willey, R. L. et al. Proc. natn. Acad. Sci. U.S.A. 83, 5038-5042 (1986).

- 14. Starcich, B. R. et al. Cell 45, 637-648 (1986). 15. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. Cell 40, 9-17 (1985).
- Sanchez-Pescador, R. et al. Science 227, 484-492 (1985).
- Ratner, L. et al. Nature 313, 227-284 (1985).
 Muesing, M. A. et al. Nature 313, 450-458 (1985).
 Clavel, F. et al. Science 233, 343-346 (1986).

- 20. Brun-Vézinet, F. et al. Lancet i, 128-132 (1987) 21. Clavel, F. et al. N. Engl. J. Med. (in the press).

- Clavel, F. et al. Nature 324, 691-695 (1986).
 Daniel, M. D. et al. Science 228, 1199-1204 (1985).
 Kanki, P. J., Alroy, J. & Essex, M. Science 230, 951-954 (1985).
- Benveniste, R. E. et al. J. Virol. 60, 483-490 (1986).
 Murphey-Corb, M. et al. Nature 321, 435-437 (1986).
- 27. Sonigo, P. et al. Cell 42, 369-382 (1985).
- Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. Science 229, 69-73 (1985).
 Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W. Science 229, 74-77 (1985).
- 30. Sodroski, J. et al. Nature 321, 412-417 (1986)
- 31. Feinberg, M. B., Jarret, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. Cell 46, 807-817 (1986)
- 32. Jones, K., Kadonaga, J., Luciw, P. & Tjian, R. Science 232, 755-759 (1986).

have a tropism for CD4-bearing cells and have elements transactivating the expression of viral genes acting at the LTR level. The evolutionary potential of these viruses is therefore striking, and we must ask whether other HIVs can emerge as long as a favourable epidemiological situation is provided. We must take advantage of the precise delineation of the conserved structures to understand their molecular biology and develop new therapeutic tools, especially immunoprophylactics.

We thank Drs S. Aaronson, F. Brun-Vézinet, R. Desrosiers, P. Fultz, H. Marquardt and M. Muesing for exchange of scientific information; Drs S. Wain-Hobson, M. L. Michel and A. Louise for providing plasmids; Bernard Caudron for help in preparation of the computer printouts; Dr. Lautaro Perez for discussion and encouragement. M.E. is a fellow of the Leukemia Society of America, Inc.

- 33. Muesing, M., Smith, D. H. & Capon, D. J. Cell 48, 691-701 (1987).
- Rabson, A. B. et al. Science 229, 1388-1390 (1985)
 Lightfoot, M. et al. J. Virol. 60, 771-775 (1986).
- 36.
- Allan, J. S. et al. Science 228, 1091-1094 (1985)
- 37. Lee, T. H. et al. Science 231, 1546-1549 (1986). 38. Kan, N. C. et al. Science 231, 1553-1555 (1986).
- Allan, J. S. et al. Science 230, 810-813 (1985)
- Miller, J., Mac Laehlan, A. & Klug, A. EMBO J. 4, 1609-1614 (1985). 40
 - 41. Dayton, A., Sodroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. Cell 44, 941-947 (1986).
 - Hess, J. L., Pyper, J. M. & Clements, J. E. J. Virol. 60, 385-393 (1986).
 Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I. & Howard, B. H. Proc. natn.

 - Acad. Sci. U.S.A. 79, 6777-6781 (1982). Adachi, A. et al. J. Virol. 59, 284-291 (1986). 44
 - Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. Cell 41, 813-823 (1985). 45.
 - 46
 - Kreiss, J. K. et al. Nature 321, 10 (1986).
 Kreiss, J. K. et al. N. Engl. J. Med. 314, 414-418 (1986).
 Kanki, P. J. et al. Science 232, 238-243 (1986). 47.
 - 49
 - Barin, F. et al. Lancet ii, 1387-1389 (1985) 50.
 - Dalgleish, A. G. et al. Nature 312, 763-767 (1984) Klatzmann, D. et al. Nature 312, 767-768 (1984) 51.
 - McDougal, J. S. et al. Science 231, 382-385 (1985) 52
 - Maddon, P. J. et al. Cell 47, 333-348 (1986). Weiss, R. A. et al. Nature 316, 69-72 (1985) 53.
 - 54
 - 55 Clavel, F., Klatzman, D. & Montagnier, L. Lancet i, 879-880 (1985).
 - Lasky, L. A. et al. Science 223, 209-212 (1985). 56
 - Messing, J. & Viera, J. Gene 19, 269-276 (1982).
 - Sanger, F., Nicklen, S. & Coulson, A. R. Proc. natn. Acad. Sci. U.S.A. 74, 5463-5467 (1977). Gorman, C. M., Motfat, L. F. & Howard, B. H. Molec. Cell. Biol. 2, 1044-1051 (1982). 58 59.
 - Graham, F. L. & Van der Eb, A. J. Virology 52, 456-467 (1973). 60

 - Wilbur, W. J. & Lipman, D. J. Proc. natn. Acad. Sci. U.S.A. 80, 726–730 (1983).
 Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. Proc. natn. Acad. Sci. U.S.A. 80, 3618–3622 (1983).
 - 63. Kiyokawa, T., Yoshikura, H., Hattori, S., Seiki, M. & Yoshida, M. Proc. natn. Acad. Sci. U.S.A. 81, 6202-6206 (1984).
 - 64. Sagata, N. et al. Proc. natn. Acad. Sci. U.S.A. 82, 677-681 (1985).
 - 65. Schwartz, D. E., Tizard, R. & Gilbert, W. Cell 32, 853-869 (1983)

I FTTERS TO NATURE

Switching phenomena in a new 90-K superconductor

P. H. Hor*, R. L. Meng*, C. W. Chu*, M. K. Wu†, E. Zirngiebl[‡], J. D. Thompson[‡] & C. Y. Huang[§]

- * Department of Physics, University of Houston, Houston, Texas 77004, USA
- † Department of Physics, University of Alabama, Huntsville, Alabama 35899, USA
- [‡] Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

§ Lockheed Palo Alto Research Laboratory, Palo Alto, California 94304-1191, USA

Recently, Wu et al.¹ and Hor et al.² have shown that $Y_{1,2}Ba_{0,8}CuO_{4-\delta}$ is a superconductor with a superconducting onset temperature at ~92 K as determined by their resistivity and a.c. susceptibility measurements. Because the magnetic properties are important in describing the nature of superconductivity, we have measured the d.c. magnetic moment of this material. Here we show that this material cooled in zero field or in a high field (H_{cool})

90 G) is diamagnetic below $T_{\rm cm} \approx$ 90 K, consistent with the previous measurements^{1,2}. However, when the sample is cooled in a small field (≤ 85 G), the magnetization, M, first becomes negative (diamagnetic) below T_{cm} , but further cooling results in a jump of M to a positive value at low temperature. We have also observed this switching by the application of an additional small field when the sample was cooled in a small field.

The $Y_{1.2}Ba_{0.8}CuO_{4-\delta}$ sample was prepared as described in ref. 1. The X-ray diffractograms reveal that the sample has multiple phases, devoid of the K₂NiF₄ structure. From the electrical resistance measurement, the superconducting onset temperature is $T_{co} \approx 94.5$ K and the resistance becomes 'zero' below $T_0 = 92$ K indicating that the sample is a superconductor with a rather narrow transition width. A Quantum Design superconducting quantum interference device (SQUID) magnetometer has been employed to measure the magnetization of the sample as a function of temperature and magnetic field. When the sample is cooled under zero field conditions, we have found that M is diamagnetic below T_{cm} and the susceptibility below ~25 K reaches ~35% of that of perfect diamagnetism $(-1/4\pi).$

We have also measured M when the sample is cooled in a field, H_{cool} . In Fig. 1, the magnetization obtained at various