Journal of Virology, Aug. 1993, p. 4639–4650 Vol. 67, No. 8
0022-538X/93/084639-12$02.00/0
Copyright © 1993, American Society for Microbiology

Analysis of Human Immunodeficiency Virus Type 1 *nef* Gene Sequences Present In Vivo

DIANE C. SHUGARS,[1,2] MARILYN S. SMITH,[3]† DEBORAH H. GLUECK,[4]
PASCALE V. NANTERMET,[3] FRANÇOISE SEILLIER-MOISEIWITSCH,[4]
AND RONALD SWANSTROM[3,5]*

Departments of Microbiology and Immunology,[1] Biostatistics,[4] and Biochemistry and Biophysics,[5] UNC School of Dentistry,[2] and UNC Lineberger Cancer Research Center,[3] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Received 26 January 1993/Accepted 29 April 1993

The *nef* genes of the human immunodeficiency viruses type 1 and 2 (HIV-1 and HIV-2) and the related simian immunodeficiency viruses (SIVs) encode a protein (Nef) whose role in virus replication and cytopathicity remains uncertain. As an attempt to elucidate the function of *nef*, we characterized the nucleotide and corresponding protein sequences of naturally occurring *nef* genes obtained from several HIV-1-infected individuals. A consensus Nef sequence was derived and used to identify several features that were highly conserved among the Nef sequences. These features included a nearly invariant myristylation signal, regions of sequence polymorphism and variable duplication, a region with an acidic charge, a $(Pxx)_4$ repeat sequence, and a potential protein kinase C phosphorylation site. Clustering of premature stop codons at position 124 was noted in 6 of the 54 Nef sequences. Further analysis revealed four stretches of residues that were highly conserved not only among the patient-derived HIV-1 Nef sequences, but also among the Nef sequences of HIV-2 and the SIVs, suggesting that Nef proteins expressed by these retroviruses are functionally equivalent. The "Nef-defining" sequences were used to evaluate the sequence alignments of known proteins reported to share sequence similarity with Nef sequences and to conduct additional computer-based searches for similar protein sequences. A gene encoding the consensus Nef sequence was also generated. This gene encodes a full-length Nef protein that should be a valuable tool in further studies of Nef function.

Sequence variability is a well-documented feature of RNA viruses in general and of the human immunodeficiency viruses (HIVs) in particular. Sequence variability arises from an error-prone viral polymerase, recombination during virus replication, and selective pressures exerted by host immunosurveillance. Sequence heterogeneity has been noted among virus isolates obtained from the same individual (14, 19, 52), related contacts (4, 8), and different unrelated individuals (35, 47). The degree of sequence heterogeneity also varies among the different HIV genes (40, 47), with *env* being the most diverse (26). Sequence diversity among virus isolates may result in altered biological properties such as growth kinetics, virulence, cellular tropism, and interactions with the immune response.

Most efforts to characterize the sequence variation of HIV isolates have been directed toward the structural gene *env* (4, 8, 26, 35, 42). Few studies have evaluated the sequence variability of nonstructural, auxiliary genes such as *nef*. Extensive sequence polymorphism has been demonstrated among *nef* genes obtained from several cultured isolates (47, 50). Because sequence changes occur during the propagation of isolates in culture (43), the sequence variation present in these cultured viruses may not accurately reflect the sequence heterogeneity present in vivo. Reports of sequence diversity in uncultured HIV type 1 (HIV-1) *nef* genes have been limited to the sequences of *nef* genes obtained from sequential isolates of one infected individual (14) and to the sequences of six *nef* genes derived from three infected individuals (27). An extensive analysis of Nef sequence variation among isolates derived from a larger sampling of unrelated individuals has not been described.

The *nef* gene is located near the extreme 3′ end of the HIV-1 genome and overlaps the 3′ long terminal repeat (LTR) (2). A similar gene is found in all primate lentiviruses but is absent in other retroviruses. The Nef protein is myristylated (2, 20, 24) and is localized predominantly to the cytoplasm and cytoplasmic membranes of infected cells (20, 25), although *nef* expression has also been detected within the nucleus (33). Deletion studies have shown that *nef* is dispensable for virus replication in cultured cells (38, 60). However, *nef* expression is essential in vivo for the maintenance of high virus loads and disease progression in macaques infected with the related simian immunodeficiency virus (SIV) (32). SIV *nef* mutants appear competent to carry out some replication in vivo, as shown by the induction of a protective immune response within infected macaques (12). Conservation of this gene among HIV-1, HIV-2, and the SIVs and the predominance of *nef*-specific transcripts early after infection (46, 51) further argue for a vital role for Nef in the virus life cycle.

Three major approaches have been taken to explore the function of *nef*. In the first approach, the growth rate of viruses containing various *nef* mutations has been evaluated in cultured cells (reviewed in references 11 and 28). Initial reports suggested that *nef* acts as a negative regulator by repressing LTR-mediated transcription and thus potentially contributing to the maintenance of viral latency in vivo. Although results from some studies have supported this model, other studies have observed either no effect, a positive effect, or differential effects of *nef* expression on replication rates, depending on factors such as the isolate

* Corresponding author.
† Present address: Département de microbiologie et immunologie, Université de Montréal, Montréal, Québec H3C 3J7, Canada.

tested, the disease state of the donor, and the multiplicity of infection (reviewed in reference 28).

A second approach has been to investigate the effects of *nef* expression on host cell functions. Nef has been reported to down-regulate the expression of three cellular proteins: CD4 (21, 21a, 24, 61), interleukin-2 (39), and NF-κB (48). These cellular proteins are intimately involved in the initiation and maintenance of an effective immune response. Therefore, Nef may allow infected cells to remain undetected by immunosurveillance, perhaps by blocking the activation of infected lymphocytes (21a). However, conflicting reports (reviewed in reference 28) underscore the need to evaluate further the effects of *nef* expression on these cellular activities.

A third approach has involved the identification of known cellular or viral proteins that share regions of amino acid sequence similarity with Nef. This approach is based on the observation that proteins sharing a certain functional property also share residues that are critical for function. From searches of protein sequence library data bases, regions of sequence similarities with the following proteins have been proposed: guanine-binding regulatory (G) proteins (24, 25), ATP-specific protein kinases (54), Bel3 protein of the human spumaretrovirus (41), class II human leukocyte antigens (HLAs) (64), interleukin-2 receptor (54), thyrotropin receptor (7), non-DNA-binding leucine zipper transcription factors (53), and scorpion peptides (65). Because these proteins are quite diverse in their cellular activities, subcellular localization, and amino acid sequences, it is doubtful that Nef shares functional properties with all of these proteins. The lack of a biochemical assay for Nef function in the context of virus replication, however, precludes the direct evaluation of the proposed sequence alignments for biological significance.

The purpose of this study was to characterize the sequence variability of HIV-1 *nef* genes obtained directly from infected individuals. We hypothesized that the encoded amino acids critical to Nef function are highly conserved among the *nef* genes of isolates obtained from unrelated individuals. Alignment of the deduced Nef protein sequences allows these conserved amino acids to be identified. Four stretches of highly conserved amino acids, or "Nef-defining" sequences, have been identified that are present within the Nef sequences of HIV-1, HIV-2, and the SIVs. Using these Nef-defining sequences, we reevaluated the reported protein sequence comparisons and conducted additional computer-based searches for similar protein sequences. Also, a consensus amino acid Nef sequence was deduced from the uncultured, patient-derived *nef* genes. The identification of this consensus sequence represents a logical approach to delineate a functional Nef sequence. A gene encoding the consensus Nef sequence was constructed, and it encodes a full-length Nef protein as demonstrated by immunoblot analysis.

MATERIALS AND METHODS

Study population. Heparinized peripheral blood was collected from 12 male HIV-1-infected individuals who were participants in the UNC AIDS Clinical Trials Group at the University of North Carolina Hospitals (Chapel Hill, N.C.). At the time of sample procurement, all study participants had low CD4 counts (10 to 190 $CD4^+$ cells per ml of blood), displayed clinical symptoms of AIDS, and were undergoing 5′-azido-3′-deoxythymidine (AZT) therapy.

DNA preparation and nested-primer PCR amplification. Peripheral blood mononuclear cells (PBMCs) were separated by centrifugation through Lymphocyte Separation Medium (Organon-Teknika-Cappel) and lysed in lysis buffer (10 mM Tris [pH 7.5], 30 mM NaCl, 20 mM EDTA) containing 0.5% sodium dodecyl sulfate (SDS) and 0.5 mg of proteinase K per ml. Total cellular DNA was extracted with phenol, ethanol precipitated, and resuspended in TE buffer (10 mM Tris-HCl [pH 7.5], 0.1 mM EDTA).

Two sets of amplification primers were designed to anneal to highly conserved segments flanking the HIV-1 *nef* gene (47). The OUTER primers (5′ OUTER primer, 5′ AAT-AGA-GTT-AGG-CAG-GGA-TA 3′; 3′ OUTER primer, 5′ CTG-GTC-TAA-CCA-GAG-AGA-CCC-AGT-AC 3′) annealed to positions 8338 to 8358 and 9533 to 9558, respectively, of the HIV-1 HXB2 sequence (49). The INNER primers (5′ INNER primer, 5′ CTC-GCA-GTC-TAG-AAG-AAT-AAG-ACA-GGG-CTT-GGA-AAG-G 3′; 3′ INNER primer, 5′ CGT-CCA-GAA-TTC-GGA-AAG-TCC-CCA-GCG-GAA-AGT-C 3′) annealed to positions 8754 to 8782 and 9436 to 9457 and included restriction sites for *Xba*I and *Eco*RI, respectively (underlined).

Because HIV-1 proviral DNA is often present at low copy numbers, we subjected total cellular DNAs to two rounds of nested-primer polymerase chain reaction (PCR) amplification (57) to obtain sufficient quantities for cloning. The first amplification round was performed in 30 cycles with the OUTER primers (1 μM each), 0.5 μg of DNA, 200 μM each of the four deoxyribonucleoside triphosphates, 10 mM Tris-HCl (pH 7.5), 1.5 mM $MgCl_2$, and 2.5 U of AmpliTaq polymerase (Perkin-Elmer Cetus). Each cycle consisted of denaturation for 1 min at 94°C, annealing for 1 min at 55°C, and elongation for 1 min at 72°C, with the denaturation time lengthened to 5 min during the first cycle and the elongation time lengthened to 10 min during the last cycle. The PCR product from the first amplification round (approximately 1,200 bp long) was diluted either 1:10 or 1:100 in fresh PCR mixture containing the INNER primers and subjected to an additional 30 cycles of amplification. PCR products from the second amplification (approximately 730 bp long) were visualized by ethidium bromide staining after agarose gel electrophoresis. Control reactions included DNA from uninfected CEM cells (a gift from M. Cloyd) and CEM cells persistently infected with the HIV-1 HXB2 isolate (49).

Cloning and sequencing. PCR products from the second amplification were gel purified, and the DNAs were eluted, cleaved with *Xba*I and *Eco*RI, and cloned by standard techniques. Approximately 20% of the *nef*-containing clones were selected for DNA sequence determination with Sequenase (United States Biochemical) by the dideoxy chain termination method (55). To reduce heteroduplex formation from reannealing during the PCR procedure, we used the most dilute PCR product per sample that gave a detectable band after the second amplification. To improve sampling, PCR products were generated from two or more independent amplifications prior to cloning.

Nucleotide misincorporation and recombination rates of *Taq* polymerase. PCR-induced errors resulting from nucleotide misincorporation and recombination are a significant concern when sequencing PCR-derived products. To determine the nucleotide misincorporation rate of *Taq* DNA polymerase in our amplification conditions, we mixed cloned HIV-1 HXB2 DNA (approximately 100 *nef* gene copies) with uninfected CEM DNA (0.5 μg) and subjected the DNAs to nested-primer PCR amplification. Eleven of the amplified HXB2 *nef* genes were then cloned and completely sequenced. Of the 6,831 bases sequenced, only 4 base changes

from the known HXB2 sequence were detected, for a misincorporation rate of 0.59 base change per 1,000 bases sequenced, or 0.37 base change per *nef* gene amplified. This misincorporation rate of approximately 10^{-5} nucleotides per cycle is comparable to the previously reported *Taq* DNA polymerase misincorporation rates (18, 23). All of the base changes involved G-to-A transitions, and no hot spot for nucleotide misincorporation was noted. For comparison, we determined the frequency of nucleotide substitutions of the seven *nef* genes derived from Pt357 relative to the consensus sequence derived for that patient. Of the 4,582 bases sequenced, we detected 154 base changes, or 33.6 base changes per 1,000 bases sequenced.

Sequence error may also arise from PCR-induced hybrid sequences generated by recombination, the switching of a partially extended, denatured amplification primer from one template molecule to another during annealing. To measure the recombination rate under our PCR conditions, we mixed HIV-1 NL432 DNA (1) (approximately 65 *nef* gene copies) and HIV-1 HXB2 DNA (approximately 35 *nef* gene copies) with 0.5 μg of uninfected CEM DNA. The *nef* genes were then amplified by nested-primer PCR amplification. We determined the DNA sequences of the 5'- and 3'-terminal 100 nucleotides of 95 amplified *nef* genes. Sixty sequences corresponded to the NL432 DNA sequence, while the remaining 35 sequences corresponded to the HXB2 DNA sequence. Recombination was detected in seven clones, or 7.4% of the amplified sequences. This recombination frequency is similar to the recombination frequency of approximately 5% previously described for *tat* (43). Since the output DNA sequence ratio of 1.7 to 1 (NL432 to HXB2) approximated the input DNA ratio of 2 to 1 (NL432 to HXB2), the mix of *nef* sequences obtained after PCR amplification likely reflects the mix of sequences present in the PBMC DNAs.

Sequence analyses and protein library data base searches. Nucleotide and amino acid sequences were aligned by the PileUp program of the University of Wisconsin Genetics Computer Group sequence analysis software program (GCG) (16). Additional sequence comparisons were conducted by the BestFit, FastA, TFastA, Gap, and WordSearch programs of GCG. To search for protein sequences that share sequence similarity with Nef proteins, we searched the PIR and SwissProtein library data bases with query patterns based on the four Nef-defining sequences. These searches were conducted by the GCG WordSearch, TFastA, and FindPatterns programs and the Analysepl program developed by Staden (59). Nef sequences were searched for sequence and structural motifs by the Motifs and ProfileScan programs of GCG, respectively.

Cells and retroviral vectors. HeLa cells and the amphotrophic retroviral packaging cell line PA317 (44) were grown in Dulbecco's modified Eagle's medium containing glucose (4.5 g/liter), 2 mM L-glutamine, and 10% fetal calf serum. From the aligned 54 Nef protein sequences, we derived an HIV-1 consensus Nef sequence representing the predominant amino acid at each position. DNA encoding this consensus Nef sequence was generated by a combination of overlapping-primer PCR amplification (17) and site-directed mutagenesis in which unique regions of four patient-derived *nef* genes were coamplified into a full-length, synthetic consensus *nef* gene (details to be described elsewhere). After the sequence was confirmed, the consensus *nef* gene was cloned into the Moloney murine leukemia virus-based retroviral vector pLXSN (45) at the *Eco*RI site in the forward (pLConsNefSN) and reverse (pLConsFenSN) orientations. Plasmid DNA was purified by cesium chloride gradient centrifugation. DNA concentration was determined by A_{260}. The molecule clone containing the HIV-1 consensus *nef* gene will be deposited in the National Institutes of Health AIDS Repository.

Generation of HeLa cells stably expressing the consensus Nef protein. Retrovirus-mediated gene transfer was used to establish HeLa cells that stably express the consensus Nef protein. PA317 cells (5×10^5 per 60-cm dish) were transiently transfected with 10 μg of pLConsNefSN or pLConsFenSN by the calcium phosphate method (45). Virus-containing supernatants were harvested 48 h later and stored at −70°C. HeLa cells (10^6 per 100-cm dish) were either mock infected or infected with 1 ml of the appropriate virus supernatant for 2 h at 37°C, washed once with phosphate-buffered saline (PBS), and then incubated with fresh medium at 37°C. Twenty-four hours after infection, cells were placed under G418 selection (0.5 mg of active G418 per ml, Geneticin; GIBCO-BRL). G418-resistant colonies were pooled and maintained under G418 selection for approximately 2 months; then G418 selection was removed.

Immunoblot analysis. Pooled G418-resistant HeLa cells were assayed for expression of the consensus Nef protein following removal of G418 selection. Mock-infected HeLa cells or HeLa cells infected with the pLConsNefSN or pLConsFenSN retroviruses were washed with PBS, trypsinized, and lysed in RIPA-PI buffer (0.02 M Tris base [pH 8.0], 2 mM EDTA, 0.15 M NaCl, 1% Triton X-100, 1% deoxycholate, 2 μg of aprotonin per ml, 1 μg of pepstatin A per ml, 100 μg of phenylmethylsulfonyl fluoride per ml, 2 μg of leupeptin per ml). Cellular lysates were clarified by brief centrifugation and electrophoresed through an SDS-13% polyacrylamide gel. The resolved proteins were then transferred onto a nitrocellulose membrane, blocked with 3% gelatin in TBST buffer (50 mM Tris-HCl [pH 6.8], 0.2 M NaCl, 0.05% Tween 20), and incubated with rabbit anti-Nef antiserum (a gift from L. Ratner, Washington University, St. Louis, Mo.) overnight at 4°C. The blot was reacted with alkaline phosphatase-conjugated goat anti-rabbit immunoglobulin G (Cappel), and color was developed with the substrates nitroblue tetrazolium (Promega) and 5-bromo-4-chloro-3-indolylphosphate (Promega) in alkaline phosphatase buffer (100 mM Tris [pH 9.5], 100 mM NaCl, 5 mM $MgCl_2$).

Nucleotide sequence accession numbers. Nucleotide and amino acid sequences have been submitted to GenBank and have been assigned accession numbers L15476 to L15529.

RESULTS AND DISCUSSION

Characteristics of the patient population. The nucleotide sequences from 54 unique *nef* genes were determined from uncultured PBMC DNAs isolated from 12 unrelated infected individuals. These 54 unique genes were obtained from a total of 56 sequenced *nef* genes. Table 1 summarizes the characteristics of the patient population and the number of genes per patient represented in the analysis. All patients had advanced HIV disease, as documented by low CD4 counts (≤190 counts per ml of blood) and AIDS-defining clinical symptoms.

Intraindividual and interindividual sequence variations. Pairwise comparisons of nucleotide and predicted amino acid sequences revealed extensive variation between the 54 *nef* genes (Table 1). In general, the sequence variation noted among *nef* genes obtained from the same individual (intraindividual variation) was generally less than the sequence

TABLE 1. Description of patient population for analysis of *nef* sequence variation

Sample (n = 12)	CD4 (count/ml)[a]	No. of genes sequenced (n = 54)	Pairwise comparison (intraindividual % difference[b])	
			Nucleotides[c]	Amino acids[d]
Pt357	140	7	1.2–11.2	2.3–18.5
Pt179	30	6	0.8–10.9	2.4–20.2
Pt102	150	6	1.1–3.5	2.4–5.3
Pt164	10	6	0.6–2.3	1.0–3.4
Pt227	10	5	2.9–8.9	1.9–16.6
Pt248	40	5	1.0–8.9	1.5–14.6
Pt166	100	5	1.1–9.8	1.0–15.8
Pt192	20	4	2.2–4.0	2.3–5.1
Pt233	190	3	3.2–10.3	4.9–16.5
Pt171	110	3	0.5–4.0	0.5–4.4
Pt226	80	2	ND[e]	ND
Pt175	20	2	ND	ND

[a] The CD4 count per milliliter of blood at the time of sample collection.
[b] Percent difference was calculated as the number of sequence substitutions per number of possible positions × 100.
[c] Interindividual, 0.3 to 13.2%.
[d] Interindividual, 0.5 to 22.7%.
[e] ND, not determined because of small sample size.

variation found among *nef* genes obtained between different individuals (interindividual variation). At the nucleotide level, the intraindividual variation ranged from 0.6 to 2.3% (Pt164) to 1.2 to 11.2% (Pt357), and the interindividual variation was 0.3 to 13.2%. At the amino acid level, the intraindividual variation ranged from 0.5 to 20.2%, while the interindividual variation was 0.5 to 22.7%.

Pattern of amino acid conservation among the Nef sequences. The 54 predicted Nef amino acid sequences were optimally aligned, and a consensus Nef sequence was derived (Fig. 1). Only the nucleotide sequences that predicted unique Nef protein sequences are presented in this study.

Taq-induced sequencing errors contributed minimally to the sequence heterogeneity noted among the *nef* genes, as shown by control experiments measuring the nucleotide misincorporation and recombination rates generated by *Taq* under our PCR amplification conditions. Also, our experimental design avoided selection for particular *nef* sequences during culturing. Overall, the extent of sequence variation present within the *nef* genes is predominantly due to the naturally occurring diversity.

The pattern of sequence variation was not distributed equally throughout the length of the Nef protein. Rather, some regions included amino acids that were invariant or highly conserved among all Nef sequences (e.g., positions 72 to 78, numbered according to the consensus Nef sequence), while other regions (e.g., positions 8 to 15) included poorly conserved amino acids and in-frame deletions and/or insertions. These clusters of poorly conserved regions tended to be near the amino and carboxyl termini, as has been previously suggested (14, 27), but some variable positions also appeared internally.

Features of HIV-1 Nef sequences. Conserved features present at the N terminus of the Nef sequences included an intact initiator methionine (ATG codon) and a highly conserved myristylation signal (residues 2 to 7). Conservation of the myristylation signal was expected since the subcellular targeting of Nef proteins to cytoplasmic membranes depends on the presence of an intact myristylation signal (66). The amino acid sequence encoding the myristylation signal was nearly invariant among our HIV-1 Nef protein sequences, although considerable sequence variation has been described in the myristylation signals of Nef sequences obtained from cultured HIV-1, HIV-2, and SIV isolates (47). Except for the required glycine at position 2, much sequence diversity is allowed among functional myristylation signals encoded by a variety of membrane-bound proteins, including other myristylated retroviral proteins such as Gag (30). Therefore, the highly conserved sequence of the Nef myristylation signal suggests that this stretch of amino acids may serve a critical, yet undetermined function other than just myristylation.

The myristylation signal was followed by a stretch of seven or eight amino acids that exhibits extensive sequence polymorphism compared with other regions of the Nef sequence. Numerous nonconservative and conservative amino acid substitutions were present within this region, with the exception of a highly conserved glycine or glutamic acid at position 12 and a highly conserved tryptophan at position 13. Sequence polymorphism in this region appeared to be a shared feature of Nef protein sequences, since highly variable sequences were present in the *nef* sequences from all DNA samples tested. The pattern of amino acid variation at this site suggests a lack of selection, rather than a tolerance of specific amino acid substitutions. Therefore, this region may serve as a flexible spacer region and is predicted to lie on the external surface of the folded protein.

The methionine at position 20 has been shown to serve as a site for internal initiation, resulting in a nonmyristylated and truncated 25-kDa Nef protein rather than the full-length, myristylated 27-kDa Nef protein (29, 36, 67). The methionine at this position was highly conserved among our HIV-1 Nef protein sequences. In contrast, none of the reported SIV Nef sequences contained a methionine at this position (47), indicating either that internal initiation may not be a significant feature for generating a functional Nef protein or that other primate lentiviruses have no need for an internally initiated Nef protein.

A variably duplicated region (between positions 22 and 23) was present in 28 of the 54 Nef protein sequences derived from 8 of the 12 patients (Fig. 1). This region represented an imperfect duplication of the adjacent downstream sequence (positions 23 to 29). The duplications varied in length from 3 to 10 amino acids (data not shown). On an individual basis, the frequency of the duplication ranged from all of the genes sequenced (Pt192, Pt175, and Pt166) to one-third of the genes sequenced (Pt179 and Pt227). No duplication was noted in the sequences derived from Pt164, Pt248, Pt171, or Pt226. Similar duplications have been reported for the *nef* gene of the brain-derived HIV-1 isolates BRVA and JR-CSF (47) and from isolates obtained from circulating lymphocytes and pathologic brain tissues of infected adults and children (6). Such variable regions frequently represent loops within a folded protein structure. Because of their diversity, the amino acid sequences present at positions 7 through 23 may be a useful marker for virus typing in transmission studies.

The Nef sequences also contained the highly conserved sequence R-P-M-T-Y-K (positions 77 to 82), which is a potential recognition site for phosphorylation by protein kinase C (PKC). The PKC recognition pattern (31) can be summarized as follows: (R/K, X_{0-2})-S/T-(X_{0-2}, R/K), where the phosphoacceptor group (serine or threonine) is preceded and/or followed by a basic residue (arginine or lysine) positioned one, two, or three residues from the phosphoacceptor group (X = any amino acid). Although PKC can phosphorylate substrates containing only C-terminal or N-terminal basic residues, PKC avidly phosphorylates sub-

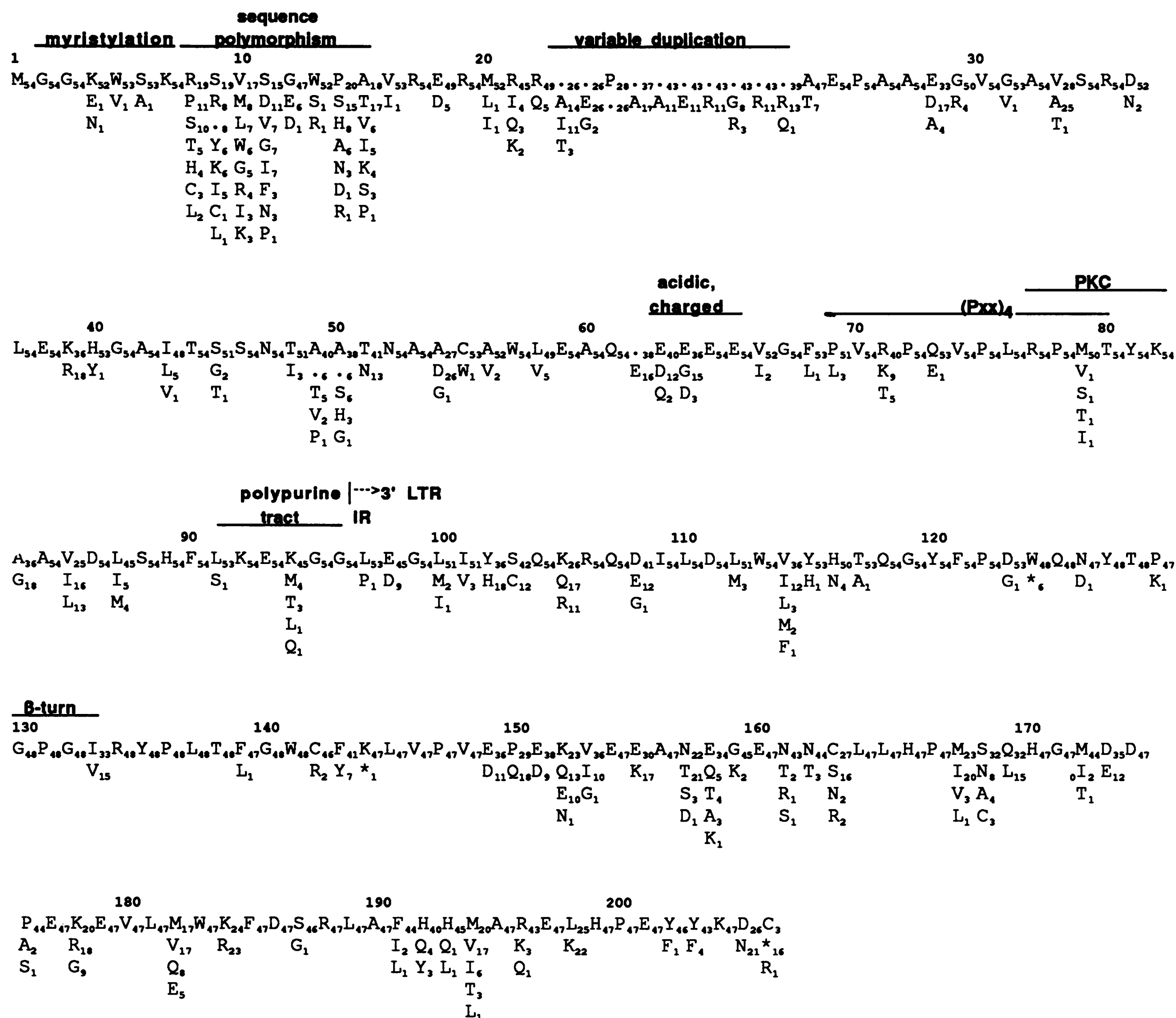

FIG. 1. Amino acid sequence variation within 54 uncultured HIV-1 *nef* genes. HIV-1 Nef protein sequences were deduced from 54 unique HIV-1 *nef* genes obtained from 12 infected individuals following nested-primer PCR amplification of uncultured PBMC DNAs. The top line depicts the consensus Nef sequence with the predominant residue at each position. The amino acid variation is represented below each position. Subscripted numbers reflect the number of times that amino acid was found at that position. Position numbering is for the consensus Nef sequence. Gaps (.) were introduced to maximize the alignment. Residues following premature termination codons (*) were not included in this summary. The locations of the myristylation signal, sequence polymorphism, variable duplication, highly charged/acidic region (acidic, charged), (Pxx)$_4$ repeat sequence, putative PKC phosphorylation site (PKC), PPT, 5' border of the 3' LTR (|--->), IR, and predicted beta-turn are shown above the consensus Nef sequence (see text for details).

strates that possess basic residues at both positions (31). All the HIV-1 Nef sequences, as well as the Nef sequences from HIV-2 and the SIVs, contained a putative PKC recognition site which included both C-terminal and N-terminal basic residues (Fig. 1 and 2). Therefore, PKC-mediated phosphorylation at this site is a potential biochemical modification of Nef proteins. Although the threonine at position 15 has been reported to be a target of PKC-catalyzed phosphorylation (24, 25), this particular potential phosphorylation site (T/S) was present in only 20 of the HIV-1 Nef sequences and was poorly conserved among the Nef protein sequences of HIV-2 and SIV (Fig. 2). Another poorly conserved potential PKC recognition site (S-X-K/R) was found at positions 103 to 105 in 37 of the HIV-1 Nef sequences. Other weakly conserved potential phosphorylation sites (31) encoded by the HIV-1 *nef* sequences included a casein kinase II recognition site [(S/T)-X-X-(D/E)] at positions 15 to 18 (20 Nef sequences) and 157 to 160 (24 Nef sequences) and a tyrosine kinase recognition site [(R/K)-X-X-X-(D/E)-X-X-X-Y] at positions 94 to 102 (36 Nef sequences).

Three other sequence features can be recognized among the Nef sequences. First, a highly conserved acidic region composed mainly of glutamic acid residues was located at positions 62 to 65. Because of its highly charged nature, this stretch of amino acids may lie on the external surface of the folded protein. Second, a sequence containing a highly

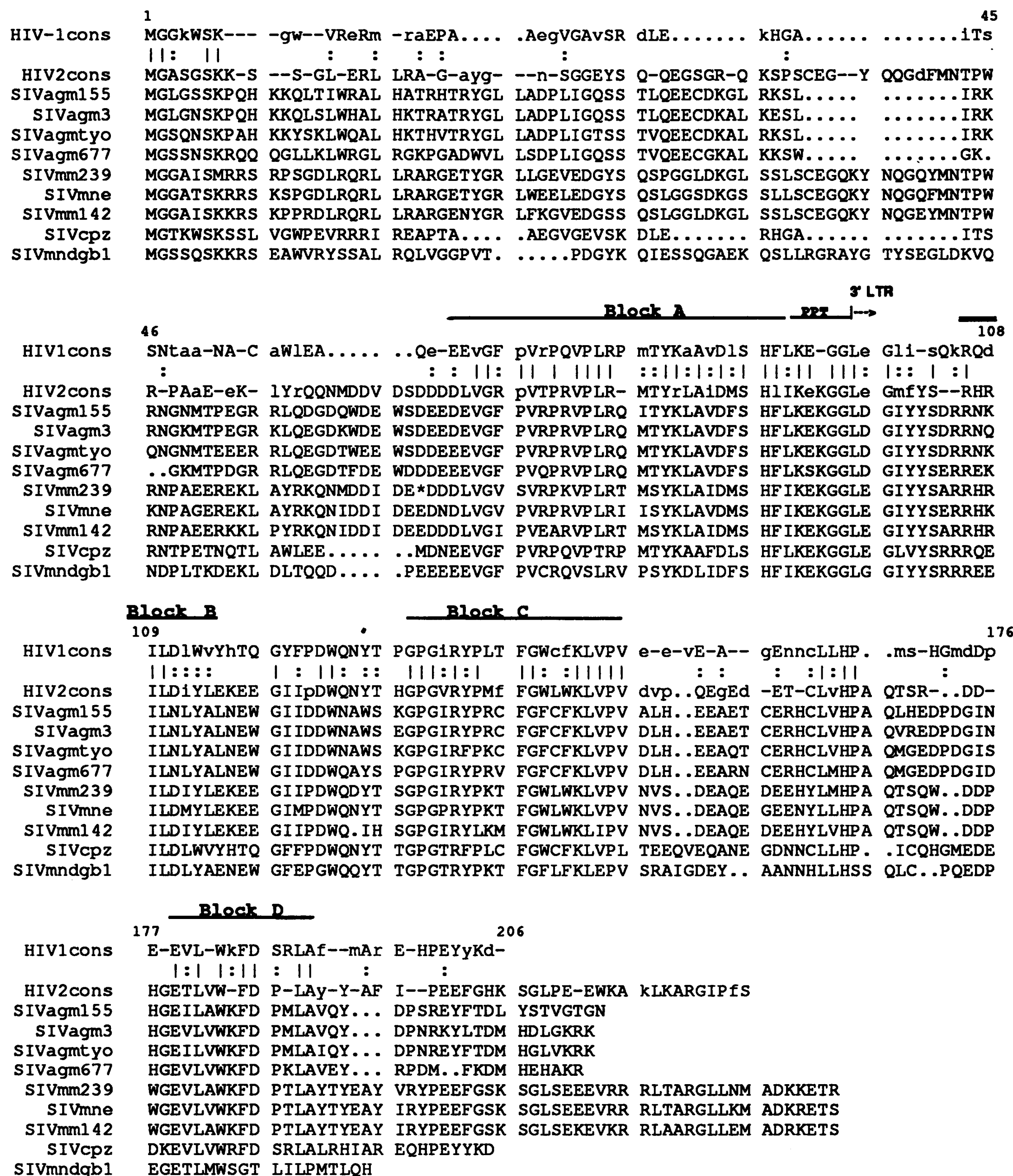

FIG. 2. Sequence similarity between invariant and highly conserved residues of the patient-derived HIV-1 consensus Nef sequence and the Nef protein sequences from HIV-2 and SIV. The consensus Nef sequence (HIV-1cons) was aligned with the HIV-2 consensus Nef protein sequence (HIV2cons; derived from HIV-2 strains ROD, GH1, D194, BEN, ISY, and ST [47]) and the Nef protein sequences from 9 SIV strains (47), using the PileUp program of GCG (16). Because of the extensive sequence heterogeneity noted among SIV Nef sequences, these sequences were not summarized into a consensus sequence. Gaps (.) were introduced to maximize the regions of alignment. HIV-1cons and HIV2cons depict the sequence variations within the 54 deduced HIV-1 Nef protein sequences and the 6 HIV-2 Nef sequences. The sequence variation at each amino acid position within HIV-1cons and HIV2cons is defined as being one of the following: 95 to 100% of the residues are identical (upper case); 95 to 100% of the residues have only conservative substitutions (lower case); less conserved (−). The amino acid sequence similarity of the HIV2cons and the SIV Nef sequences with HIV-1cons is defined as one of the following: 90 to 100% of the residues are identical to the corresponding HIV-1cons residue (|); 80 to 100% of the residues represent conservative substitutions of the corresponding HIV-1cons residue (:). Blocks A, B, C, and D indicate stretches of amino acids that are highly conserved among the Nef proteins of HIV-1, HIV-2, and the SIVs. The locations of the PPT and the 5′ border of the 3′ LTR (| →) are indicated above the alignment. Position numbering is for the HIV-1 consensus Nef sequence. Conservative amino acids are grouped as follows: A, S, T, G, and P; V, L, I, and M; F, Y, and W; R, K, and H; D, E, N, and Q; and C (13). * indicates a premature stop codon.

conserved proline followed by two residues was repeated four times in succession [$(Pxx)_4$] within positions 69 to 80. This repeat was also highly conserved among the Nef protein sequences obtained from cultured HIV-2 and SIV isolates (47), although there the Pxx sequence is repeated three times rather than four times (Fig. 2). This $(Pxx)_4$ repeat sequence can be found in numerous proteins with diverse biological properties, including the proline-rich domain of the murine leukemia virus envelope protein (56). Although the significance of this repeat sequence within the Nef protein sequence is unknown, it probably represents a structural element. Third, the highly conserved G-P-G-I/V sequence located at positions 130 to 133 is highly predictive of a beta-turn (10). The third variable domain (V3) of HIV-1 *env* encodes the similar sequence G-P-G-R which forms the tip of the V3 loop structure (35).

Conservation of *cis*-acting RNA and DNA elements. During virus replication, the U3 and U5 regions located at the ends of the retroviral RNA genome are duplicated during reverse transcription to form the LTRs (reviewed in reference 62). The polypurine tract (PPT) is one essential element for reverse transcription that is located within the *nef* coding sequence. The PPT serves as the primer for initiation of second-strand DNA synthesis. The PPT consists of a purine-rich region of 16 nucleotides that lies immediately upstream of the 5′ border of the 3′ LTR. Among our *nef* genes, amino acid sequences encoded in the PPT were highly conserved (codon positions 91 to 96; i.e., glutamic acid, lysine, and glycine residues) and reflected the purine-rich composition of this sequence. The nucleotide sequences maintained the A/G-rich composition while encoding different amino acids, exemplified by the amino acid variation noted at position 94. Therefore, sequence conservation in this region is probably due to constraints at the level of the RNA sequence for virus replication, rather than at the protein level.

The inverted repeat (IR) is a critical element of the LTR that is located within the *nef* gene. In retroviruses, the IR is a short, inverted repeat located at the 3′ end of U5 and the 5′ end of U3. IR sequences are critical for proper cleavage and integration of the proviral DNA into the host chromosome (reviewed in reference 62). Among our *nef* genes, the nucleotide sequences encoding the IR were relatively conserved (data not shown) and reflected the sequence variation previously described for cultured HIV-1 isolates (47).

Studies of HIV LTR gene expression in vitro have identified at least five potential binding sites for cellular transcription factors in the U3 region overlapping the 3′ end of the *nef* gene (reviewed in reference 22). These regions include sites for chicken ovalbumin upstream promoter transcription factor (COUP-TF), activator protein (AP-1), nuclear factor 1 of activated T cells (NFAT-1), upstream stimulatory factor (USF), and T-cell factor 1 alpha (TCF-1α). We aligned five regions within the deduced HIV-1 Nef sequences that overlap these putative sites. The locations of these sites within the consensus Nef codon sequence are as follows: COUP-TF, positions 125 to 140; AP-1, positions 133 to 135 and 138 to 139; NFAT-1, positions 151 to 163 and 176 to 180; USF, positions 190 to 195; and TCF-1α, positions 202 to 206. The nucleotide sequences that encode these binding sites were fairly conserved within the *nef* sequences but, in general, do not overlap the Nef-defining sequences (data not shown). The exceptions are the COUP-TF and AP-1 sites which partially overlap block C.

Premature stop codons at position 124. An unexpected finding was the clustering of premature stop codons at position 124 (normally encoding tryptophan) in six sequences derived from 4 of the 12 individuals. Analysis of the sequence similarity of these six sequences with the other sequences from these individuals revealed that the stop-encoding sequences were not a result of contamination with HXB2, an HIV-1 clone used in our laboratory that also encodes a premature stop at this position (data not shown). The clustering of premature stops appeared to be specific for position 124, given the lack of premature stop codons at nearby tryptophan-encoding codons (positions 113 and 141). Furthermore, the HIV-1 BH10 clone (49) and two *nef* clones isolated from uncultured PBMCs of two unrelated, HIV-1-infected individuals (27) also encode premature stops at position 124. The premature stops do not appear to be the result of G-to-A hypermutations, concentrated regions of extensive G-to-A transitions that typically involve GpA dinucleotides (63), because only one of the prematurely truncated *nef* sequences (Pt357) contained a region of potential G-to-A hypermutation. In this case, there were only five G-to-A transitions within a 100-nucleotide stretch, and only one of the transitions involved a GpA dinucleotide (data not shown). No region of G-to-A hypermutation was noted in the remaining sequences encoding either a prematurely truncated or full-length Nef protein. Although most published *nef* sequences are open, other investigators have reported the presence of premature stop codons within the coding sequences of *tat* (43), *rev* (40), *pol* (58), and *env* (4, 42) genes, as well as among *nef* genes isolated from PBMCs (14, 49) and other tissues (6). Clustering at position 124 suggests that truncated Nef proteins may be generated in vivo and may display altered properties compared with the full-length Nef protein.

Identification of four Nef-defining sequences. One of the goals of this study was to identify stretches of amino acids that were highly conserved not only among the Nef sequences derived from uncultured HIV-1 isolates, but also among HIV-2 and SIV isolates. The rationale for this goal is that, regardless of their source, all functional Nef proteins must contain certain amino acids that are critical for activity. To identify these highly conserved regions, we aligned the HIV-1 consensus sequence with a consensus sequence derived from six HIV-2 Nef protein sequences (HIV-2 isolates ROD, GH1, D194, BEN, ISY, and ST [47]) and the Nef protein sequences from 9 SIV isolates (AGM155, AGM3, AGMTYO, AGM677, MM239, MNE, MM142, CPZ, and MNDGLB1 [47]). We chose not to derive a consensus sequence for the SIV Nef sequences since they are the most diverse in amino acid composition and length, and some SIV isolates were derived from different hosts. From this alignment (shown in Fig. 2), four stretches of highly conserved or invariant amino acids were identified and are referred to as blocks A through D. Block A (positions 64 to 90) denotes a stretch of 27 amino acids that includes the acidic/charged region, the $(Pxx)_4$ repeat, and the potential PKC phosphorylation site. Block A was not extended to include the highly conserved amino acids overlapping the PPT, since sequence conservation here is probably due to constraints for virus replication (see above). The other blocks include block B at positions 106 to 114, block C at positions 130 to 148 (including sequences that may form a predicted beta-turn structure), and block D at positions 179 to 190. These blocks include most of the highly conserved stretches of amino acids. Exceptions are short stretches that include the conserved D-W-Q/N sequence (positions 123 to 125) located between blocks C and D, and the conserved L-L-H-P sequence (positions 164 to 167).

Although the nucleotide sequences encoding blocks B, C,

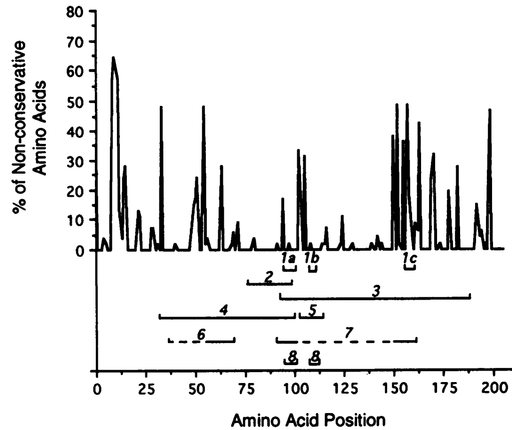

FIG. 3. Location of published protein sequence alignments in relation to the Nef amino acid variation pattern. The top portion of the graph depicts the percentage of nonconservative amino acid substitutions at each position that were derived from alignment of the 54 patient-derived Nef sequences. The percentage of nonconservative amino acids found at each position is indicated on the *y* axis; the amino acid position corresponding to the consensus Nef sequence is depicted on the *x* axis. The bottom portion of the graph shows the location along the consensus Nef sequence of the published alignments of protein sequences reported to share limited sequence similarity with various Nef proteins (7, 25, 41, 53, 54, 64, and 65). 1, G proteins: 1a = P site, 1b = C site, 1c = G site; 2, leucine zipper motif; 3, Bel3 of human spumaretrovirus; 4, HLA class II antigens; 5, interleukin-2 receptor; 6, thyrotropin receptor; 7, scorpion peptides; 8, ATP-binding site of the protein kinase family. ----, indicates variable spacing used to maximize the sequence alignments, as reported in the initial published alignments.

and D were also present within the U3 region, these three blocks did not significantly overlap any known *cis*-acting DNA element, with the exception of the AP-1 and COUP-TF binding sites which partially overlap block C. Residues that were highly conserved among the HIV-1 Nef protein sequences corresponded to highly conserved residues of Nef sequences from HIV-2 and the SIVs, making it likely that the Nef proteins from these viruses are functionally equivalent.

Reevaluation of reported sequence similarities. Several investigators have proposed sequence similarities between Nef proteins and known cellular or viral proteins based on alignments with Nef sequences deduced from cultured HIV-1, HIV-2, and SIV isolates (47). Proposed protein similarities include G proteins (24, 25), ATP-specific protein kinases (54), Bel3 protein of human spumaretrovirus (41), class II HLAs (64), interleukin-2 receptor (54), scorpion peptides (65), thyrotropin receptor (7), and non-DNA-binding leucine zipper transcription factors (53). The location of these alignments along the length of the Nef sequence is shown in Fig. 3. Because of the diverse activities displayed by these proteins, it is unlikely that Nef exhibits functional characteristics of all of these proteins. It is difficult to determine which, if any, of these diverse proteins share activity with Nef proteins. Therefore, we reassessed these reports by aligning the reported protein sequences to our HIV-1 Nef sequence data (Fig. 3) and to the four Nef-defining sequences (Fig. 4), using the alignments published in the original reports. We reasoned that the proposed regions of sequence similarity should include the Nef-defining sequences as a rigorous test of the biological significance of the alignment.

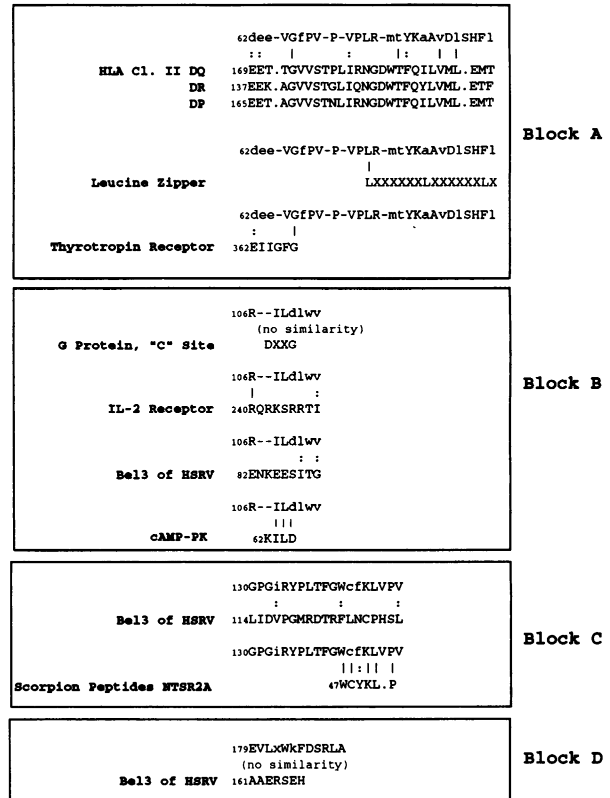

FIG. 4. Sequence similarity between Nef-defining sequences and protein sequences previously reported to share partial sequence similarity with Nef proteins. The amino acid sequences of cellular or viral proteins proposed to share sequence similarity with Nef proteins were aligned with the Nef-defining sequences, depicted as blocks A through D. These blocks represent stretches of amino acids that are highly conserved among Nef protein sequences derived from our uncultured HIV-1 isolates and from cultured HIV-2 and SIV isolates (Fig. 2). We derived consensus sequences that summarize the amino acid sequence variation present within the blocks as follows: 90% of the residues are identical (upper case); 90% of the residues represent only conservative substitutions (lower case); less conserved (−). Sequence comparisons between blocks A through D and the reported protein sequences are indicated as follows: residues are identical (|), or residues are conservative substitutions (:). Only the protein sequences that were shown in the initial published alignments and overlap the Nef-defining sequences are shown. Numbering is of the HIV-1 consensus Nef sequence. Amino acid coordinates are presented at the beginning of the sequences (see references 7, 25, 41, 53, 54, 64, and 65). Amino acid coordinates are not shown for the leucine zipper and G protein (C site) since canonical sequences are given. Gaps (.) were introduced to maximize the regions of alignment.

Alignment of the proposed protein sequences with blocks A through D (Fig. 4) revealed that portions of the eight protein sequences overlapped the Nef-defining sequences. Upon closer inspection, however, the overall extent of amino acid sequence similarity was relatively small, as demonstrated by the lack of significant sequence identify or conservation between most of the sequences.

The alignment that shared the greatest number of invariant and highly conserved amino acids was between the HLA class II antigens (DQ, DR, and DP) and block A (Fig. 4). The two sequences shared four positions of identical amino acids and four positions of highly conservative amino acids. The

region of limited sequence similarity overlapped the CD4 binding region within the immunoglobulinlike second extracellular domain of the HLA class II beta-chain (9, 34). It has been speculated that the autoimmunelike response frequently observed in patients with AIDS may be induced against the HLA class II beta-chain by the Nef protein through molecular mimicry (64). However, *nef* expressed from viral vectors (21, 21a, 24, 25) has been shown to suppress cell surface CD4 expression in human cell lines of diverse origins through an interaction that appears to involve the cytoplasmic domain of CD4 (21). These data suggest that the Nef protein and the HLA class II antigens may share a common biological property, although the sequence comparisons weakly support this hypothesis.

The ATP-binding site of protein kinases such as cyclic AMP-dependent protein kinases showed partial sequence similarity to block B. The sequence D/K-I-L-D/N, a highly conserved sequence that is found within the nucleotide binding domain of protein kinases (5), was also relatively conserved among Nef sequences (positions 108 to 111, Fig. 4). Sequences corresponding to other sites which are critical in binding and hydrolyzing ATP were not present within the Nef-defining sequences, however. These critical sequences included the glycine-rich canonical G-X-G-X-X-G sequence and a lysine positioned 15 to 28 amino acids C terminal to the glycine-rich sequence (5). Moreover, biochemical assays have failed to demonstrate ATP-binding and hydrolysis activities in purified Nef proteins (reviewed in reference 28), making it unlikely that Nef functions as an ATP-dependent protein kinase.

The remaining alignments demonstrated weak sequence relatedness between the Nef-defining sequences and reported cellular or viral proteins (Fig. 4). As an example, Guy et al. (24, 25) have reported that five Nef protein sequences (isolates HIV-1 BRU, HIV-1 ARV2, HIV-2 ROD, SIV MAC251, and SIV AGM [47]) share regions of sequence relatedness with three GTP-binding elements of G proteins, signal-transducing proteins having properties of GTP hydrolysis, GTP binding, and autophosphorylation (e.g., $p21^{ras}$). These highly conserved elements are the P site (canonical G-X-X-X-X-G-K sequence, also referred to as the A site), the C site (D-K-X-G), and the G site (N-K-X-D) (15, 37). Although weak GTP-binding and GTPase activities were exhibited by bacterially expressed and partially purified Nef proteins (24, 25), subsequent studies by other investigators have failed to demonstrate these activities in purified Nef proteins, attributing the original finding to bacterial contamination (27; reviewed in reference 28). Our sequence analysis confirms the lack of significant sequence similarity between the Nef protein sequences and G proteins. The proposed G site (positions 183 to 186) was relatively conserved among our Nef sequences, but it overlapped the LTR and did not correspond to a Nef-defining sequence. Although the C site overlapped a Nef-defining sequence (block B), the site was only partially conserved among our sequences (positions 108 to 111). Furthermore, the proposed P site (positions 94 to 99) aligned only if the chemical polarity of the amino acid sequence was inverted. Taken together, the sequence and biochemical analyses indicate that Nef fails to exhibit features suggestive of G proteins.

Searches for cellular or viral proteins that contain Nef-defining sequences. We used the most highly conserved regions in the Nef sequence to conduct data base searches for known protein sequences sharing significant similarity with Nef, using various GCG (16) and Staden (59) search programs. Our approach was to search the data bases

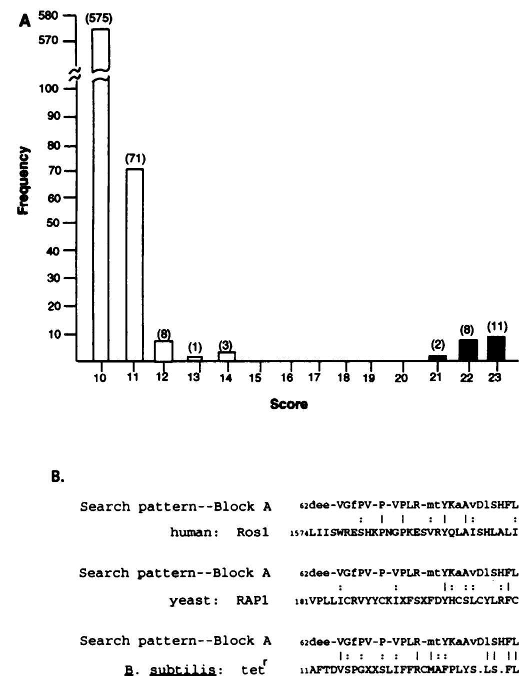

FIG. 5. Results of the protein sequence library searches for known proteins that share sequence similarity with the four Nef-defining sequences. (A) Example of results from a search of the PIR protein sequence library using the block A Nef-defining sequence. The PIR protein sequence data base library was searched by the Analysepl program (59) with the block A Nef-defining sequence E/D-E/D/N-E/D-E/L-V-G-V/F/I-P/S-V-X-R/P-Q/R/K-V-P/S-L/T-R-X-X-T/S-Y-R/K-X-A/L-X-D-X-S-H-F-L/I (abbreviated as dee-VGfPV-VPLR-mt-TKaAvDISHFL; invariant residues [upper case], highly conserved residues [lower case], poorly conserved residues [-]). The scores of the alignments (x axis) are plotted as a function of the number of matches having that score (in parentheses, y axis). Nef matches are depicted with solid bars; non-Nef matches are shown with open bars. (B) Alignments of the block A query sequence and the best non-Nef matches. The block A sequence (abbreviated as described in panel A) was aligned to the best non-Nef matches by the TFastA program of GCG (16). The best non-Nef matches were the human transmembrane tyrosine-specific kinase (ROS1) protein (two alleles, accession numbers M34353 and X51619), the *Saccharomyces cerevisiae* DNA-binding protein RAP1 (accession number P11938), and the *Bacillus subtilis* tetracycline resistance determinant (accession number X08034). Sequence similarity with block A is noted as follows: identical residues (|), highly conserved residues (:).

utilizing the Nef-defining domains (blocks A through D) individually in an initial screen. This approach readily identified all the varied Nef sequences in the data base as a distinct subset of sequences (see Fig. 5A for an example). As expected, an increasing number of matches was found with lesser match scores. The heterologous proteins with the highest scores have very different biochemical activities and align over different regions of the query sequence (Fig. 5B), indicating that most, if not all, of these alignments represent background. In an effort to avoid fortuitous alignments, we evaluated the matches having the best scores using the following criteria: (i) identify the best alignments from searches using each of the blocks individually; (ii) use this subset of sequences to find the best alignments with the other blocks; (iii) evaluate the linear order of the aligned blocks compared to the order of the blocks in Nef; and (iv)

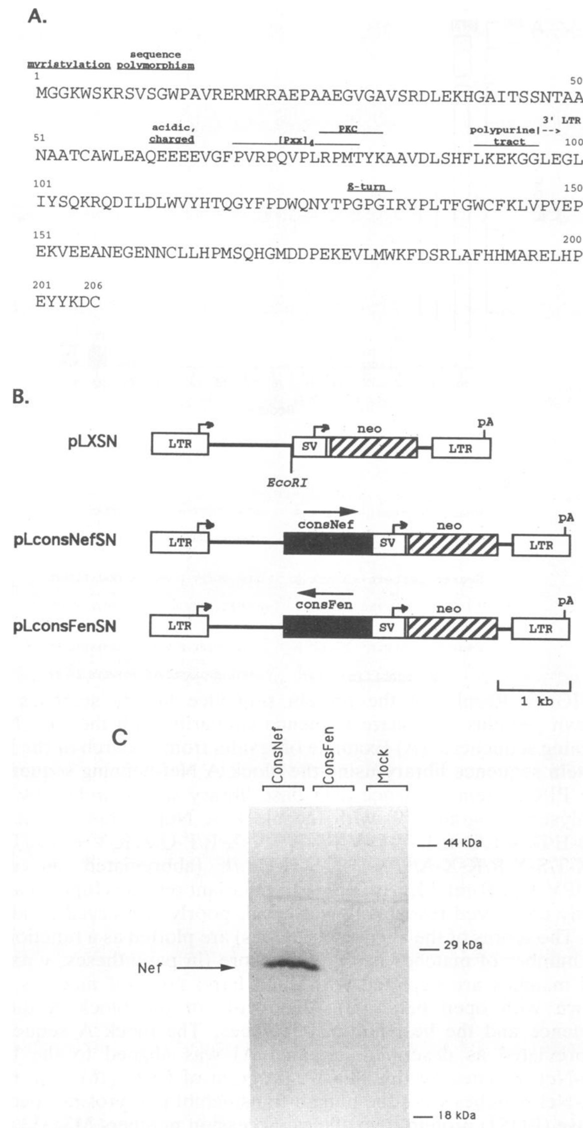

FIG. 6. Stable expression of the HIV-1 *nef* consensus gene transduced into HeLa cells. (A) Nef amino acid sequence predicted from the HIV-1 *nef* consensus gene. A gene encoding the HIV-1 consensus Nef sequence was synthesized through a combination of overlapping-primer PCR amplification and site-directed mutagenesis. The amino acid sequence of this gene corresponds exactly to the Nef consensus sequence depicted in Fig. 1 except that the proline within the region of variable duplication was omitted (see text). The locations of the myristylation signal, sequence polymorphism, highly charged/acidic region, $[Pxx]_4$ repeat, putative PKC phosphorylation site (PKC), PPT, 5′ border of the 3′ LTR (| →), and the predicted beta-turn are indicated above the sequence. Position numbering is indicated above the Nef sequence. (B) Retroviral vectors pLConsNefSN and pLConsFenSN. The HIV-1 consensus *nef* gene was cloned into the pLXSN retroviral vector at the *Eco*RI site in the forward (pLConsNefSN, consNef) and reverse (pLConsFenSN, consFen) orientations. The consensus *nef* gene is expressed from the Moloney murine leukemia virus LTR. Other abbreviations: SV, simian virus 40 early promoter; neo, neo^r gene; pA, polyadenylation site. (C) Immunoblot analysis of HeLa cells expressing the HIV-1 consensus Nef protein. PA317 cells were transiently transfected with the pLConsNefSN and pLConsFenSN retroviral DNAs. Virus was harvested and used to infect HeLa cells. G418-resistant cells were selected and analyzed via immunoblotting for expression of the HIV-1 consensus Nef protein. Cellular lysates were electrophoresed on a 13% polyacrylamide gel, transferred to nitrocellulose membrane, and probed with a rabbit anti-Nef antiserum.

assess the level of conservation of polar and charged amino acids that are invariant in Nef sequences. We reasoned that polar and charged residues are more likely to be involved in catalytic function, as opposed to hydrophobic residues that are more likely to have a structural role and thus may be less conserved. Using this approach, we were unable to identify satisfactory alignments between Nef and any heterologous protein in the available data bases. Searches for known structural and sequence motifs by the GCG ProfileScan and Motifs programs, respectively, also failed to identify known motifs within either the consensus Nef sequence or the 54 individual Nef sequences (data not shown).

The consensus *nef* gene expresses a full-length, stable Nef protein within infected HeLa cells. A gene encoding the consensus Nef amino acid sequence was constructed, using a combination of PCR amplification and site-directed mutagenesis. The sequence of the consensus *nef* gene is shown in Fig. 6A. This sequence corresponded exactly to the Nef consensus sequence depicted in Fig. 1, except that the proline at position 23 within the variable duplication region was omitted. This choice was made based on the observation that proline was present only as part of the irregular series of duplications that appeared in this region, and proline never appeared alone in this position.

We used retrovirus-mediated gene transfer to establish HeLa cells that stably expressed the consensus *nef* gene. The retroviral vectors used in this study are diagrammed in Fig. 6B. Immunoblot analysis of pooled G418-resistant infected HeLa cells demonstrated that a Nef protein of approximately 27 kDa was expressed in the cells infected with the pLConsNefSN vector, whereas no Nef protein was detected in cells infected with the control pLConsFenSN vector (Fig. 6C). In an initial examination of protein stability, the half-life of the consensus Nef protein was estimated at greater than 4 h as assessed by pulse-chase radiolabeling experiments (data not shown), comparable to what has been reported in the literature (28). Also, expression of the consensus *nef* gene in cells expressing CD4 resulted in suppression of cell surface CD4 expression (3), as has been reported previously for *nef* alleles obtained from laboratory HIV-1 strains (21, 21a, 24, 61).

The construction and expression of a *nef* gene encoding the consensus Nef sequence represents a logical approach to identify a functional *nef* gene. The lack of an assay for Nef activity in the context of virus replication makes it difficult to assess the competence of any specific *nef* allele. The consensus *nef* gene should encode an optimally functioning protein with which to explore the biological and biochemical properties of Nef.

ACKNOWLEDGMENTS

We thank Lee Ratner for providing anti-Nef antiserum, Clyde Hutchison for critical review of the manuscript, John Olsen and Victor Garcia for helpful discussions, and Lorraine Everitt for assistance with DNA sequencing. We also thank Charles van der Horst and the clinical staff of the UNC AIDS Clinical Trials Group at the University of North Carolina Hospitals for obtaining blood samples.

This work was supported by Public Health Service grants U01-AI25868, and R01-AI25321 (to R.S.) and the National Institutes of Health grant RR00046 to the University of North Carolina General

Clinical Research Center. D.C.S. was supported by a Dentist-Scientist award from the National Institute for Dental Research (K16-DE00165).

REFERENCES

1. **Adachi, A., H. E. Gendleman, S. Koenig, T. Folks, R. Willey, A. Rabson, and M. A. Martin.** 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J. Virol. **59:**284–291.
2. **Allan, J. S., J. E. Coligan, T.-H. Lee, M. F. McLane, P. J. Kanki, J. E. Groopman, and M. Essex.** 1985. A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS patients. Science **230:**810–813.
3. **Anderson, S., D. C. Shugars, R. Swanstrom, and J. V. Garcia.** 1993. Nef from primary isolates of human immunodeficiency virus type 1 suppresses surface CD4 expression in human and mouse T cells. J. Virol. **67:**4923–4931.
4. **Balfe, P., P. Simmonds, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown.** 1990. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating substitutions. J. Virol. **64:**6221–6233.
5. **Barker, W. C., and M. O. Dayhoff.** 1982. Viral *src* gene products are related to the catalytic chain of mammalian cAMP-dependent protein kinase. Proc. Natl. Acad. Sci. USA **79:**2836–2839.
6. **Blumberg, B. M., L. G. Epstein, Y. Saito, D. Chen, L. R. Sharer, and R. Anand.** 1992. Human immunodeficiency virus type 1 *nef* quasispecies in pathological tissue. J. Virol. **66:**5256–5264.
7. **Burch, H. B., E. V. Nagy, Y. G. Lukes, W. Y. Cai, L. Wartofsky, and K. D. Burman.** 1991. Nucleotide and amino acid homology between the human thyrotropin receptor and the HIV-1 Nef protein: identification and functional analysis. Biochem. Biophys. Res. Commun. **181:**498–505.
8. **Burger, H., B. Weiser, K. Flaherty, J. Gulla, P.-N. Nguyen, and R. A. Gibbs.** 1991. Evolution of human immunodeficiency virus type 1 nucleotide sequence diversity among close contacts. Proc. Natl. Acad. Sci. USA **88:**11236–11240.
9. **Cammarota, G., A. Scheirle, B. Takacs, D. M. Doran, R. Knorr, W. Bannwarth, J. Guardiola, and F. Sinigalia.** 1992. Identification of a CD4 binding site on the beta 2 domain of HLA-DR molecules. Nature (London) **356:**799–801.
10. **Chou, P. Y., and G. D. Fasman.** 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. **47:**251–276.
11. **Cullen, B. R.** 1991. Regulation of human immunodeficiency virus replication. Annu. Rev. Microbiol. **45:**219–250.
12. **Daniel, M. D., F. Kirchhoff, S. C. Czajak, P. K. Sehgal, and R. C. Desrosiers.** 1993. Protective effects of a live attenuated SIV vaccine with a deletion in the *nef* gene. Science **258:**1938–1941.
13. **Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt.** 1987. A model of evolutionary change in proteins, p. 345–352. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, 5. National Biomedical Research Foundation, Washington, D.C.
14. **Delassus, S., R. Cheynier, and S. Wain-Hobson.** 1991. Evolution of human immunodeficiency virus type 1 *nef* and long terminal repeat sequences over 4 years in vivo and in vitro. J. Virol. **65:**225–231.
15. **Dever, T. E., M. J. Glynias, and W. Merrick.** 1987. GTP binding domain: three consensus sequence elements with distinct spacing. Proc. Natl. Acad. Sci. USA **84:**1814–1818.
16. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
17. **Dillon, P. J., and C. A. Rosen.** 1990. A rapid method for the construction of synthetic genes using the polymerase chain reaction. BioTechniques **9:**299–300.
18. **Eckert, K. A., and T. A. Kunkel.** 1990. High fidelity DNA synthesis by the DNA polymerase. Nucleic Acids Res. **18:**3739–3744.
19. **Fisher, A. G., B. Ensoli, D. Looney, A. Rose, R. C. Gallo, M. S. Saag, G. M. Shaw, B. H. Hahn, and F. Wong-Staal.** 1988. Biologically diverse molecular variants within a single HIV-1 isolate. Nature (London) **334:**444–447.
20. **Franchini, G., M. Robert-Guroff, J. Ghrayeb, N. T. Chang, and F. Wong-Staal.** 1986. Cytoplasmic localization of the HTLVIII 3′ orf protein in cultured T cells. Virology **155:**593–599.
21. **Garcia, J. V., J. Alfano, and A. D. Miller.** 1993. The negative effect of human immunodeficiency virus type 1 *nef* on cell surface CD4 expression is not species specific and requires the cytoplasmic domain of CD4. J. Virol. **67:**1511–1516.
21a. **Garcia, J. V., and A. D. Miller.** 1991. Serine phosphorylation-independent downregulation of cell-surface CD4 by *nef*. Nature (London) **350:**508–511.
22. **Gaynor, R. B.** 1992. Cellular transcription factors involved in the regulation of HIV-1 gene expression. AIDS **6:**347–363.
23. **Gelfand, D. H., and T. J. White.** 1990. Thermostable DNA polymerases, p. 129–141. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols: a guide to methods and applications. Academic Press, Inc., San Diego.
24. **Guy, B., M. P. Kieny, Y. Rivière, C. Le Peuch, K. Dott, M. Girard, L. Montagnier, and J.-P. LeCocq.** 1987. F/3′ orf encodes a phosphorylated GTP-binding protein resembling an oncogene product. Nature (London) **330:**266–269.
25. **Guy, B., Y. Rivière, K. Dott, A. Regnault, and M. P. Kieny.** 1990. Mutational analysis of the HIV Nef protein. Virology **176:**413–525.
26. **Hahn, B. H., M. A. Gonda, G. M. Shaw, M. Popovic, J. Hoxie, R. C. Gallo, and F. Wong-Staal.** 1985. Genomic diversity of the AIDS virus HTLV-III: different viruses exhibit greatest divergence in their envelope genes. Proc. Natl. Acad. Sci. USA **82:**4813–4817.
27. **Harris, M., S. Hislop, P. Patsilinacos, and J. C. Neil.** 1992. In vivo derived HIV-1 *nef* gene products are heterogeneous and lack detectable nucleotide binding activity. AIDS Res. Hum. Retroviruses **5:**537–542.
28. **Hovanessian, A. G.** 1992. Third forum in virology: on the HIV *nef* gene product. Res. Virol. **143:**31–81.
29. **Kaminchik, J., N. Bashan, A. Itach, N. Sarver, M. Gorecki, and A. Panet.** 1991. Genetic characterization of human immunodeficiency virus type 1 *nef* gene product translated in vitro and expressed in mammalian cells. J. Virol. **65:**583–588.
30. **Kaplan, J. M., G. Mardon, J. M. Bishop, and H. E. Varmus.** 1988. The first seven amino acids encoded by the v-*src* oncogene act as a myristylation signal: lysine 7 is a critical determinant. Mol. Cell. Biol. **8:**2435–2441.
31. **Kennelly, P. J., and E. G. Krebs.** 1991. Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. J. Biol. Chem. **266:**15555–15558.
32. **Kestler, H. W., D. J. Ringler, K. Mori, D. L. Panicali, P. K. Sehgal, M. D. Daniel, and R. D. Desrosiers.** 1991. Importance of the *nef* gene for maintenance of high virus loads and/or development of AIDS. Cell **65:**651–662.
33. **Kienzle, N., M. Bachmann, W. E. G. Muller, and N. Muller-Lantzsch.** 1992. Expression and cellular localization of the Nef protein from human immunodeficiency virus-1 in stably transfected B-cells. Arch. Virol. **124:**123–132.
34. **König, R., L.-Y. H., and R. N. Germain.** 1992. MHC class II interaction with CD4 mediated by a region analogous to the MHC class I binding site for CD8. Nature (London) **356:**796–801.
35. **LaRosa, G. J., J. P. Davide, K. Weinhold, J. A. Waterbury, A. T. Profy, J. A. Lewis, A. J. Langlois, G. R. Dreesman, B. N. Boswell, P. Shadduck, L. H. Holley, M. Karplus, D. P. Bolognesi, T. J. Mattnews, E. A. Emini, and S. D. Putney.** 1990. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science **249:**932–935.
36. **Laurent, A. G., A. G. Hovanessian, Y. Rivière, B. Krust, A. Regnault, L. Montagnier, A. Fendeli, M. P. Kieny, and B. Guy.** 1990. Production of a nonfunctional Nef protein in human immunodeficiency virus type 1-infected CEM cells. J. Gen. Virol. **71:**2273–2281.
37. **Lochrie, M. A., and M. I. Simon.** 1988. G protein multiplicity in eukaryotic signal transduction system. Biochemistry **27:**4957–4965.
38. **Luciw, P. A., C. Cheng-Mayer, and J. A. Levy.** 1987. Mutational

analysis of the human immunodeficiency virus: the orf-B region down-regulates virus replication. Proc. Natl. Acad. Sci. USA **84**:1434–1438.

39. **Luria, S., I. Chambers, and P. Berg.** 1991. Expression of the type 1 human immunodeficiency virus Nef protein in T cells presents antigen receptor-mediated induction of interleukin 2 mRNA. Proc. Natl. Acad. Sci. USA **88**:5326–5330.
40. **Martins, L. P., N. Chenciner, B. Asjö, A. Meyerhans, and S. Wain-Hobson.** 1991. Independent fluctuation of human immunodeficiency virus type 1 *rev* and gp41 quasispecies in vivo. J. Virol. **65**:4502–4507.
41. **Maurer, B., and R. M. Flügel.** 1987. The 3'-orf protein of human immunodeficiency virus 2 shows sequence homology with the *bel3* gene of the human spumaretrovirus. FEBS Lett. **2**:286–288.
42. **McNearney, T., Z. Hornickova, R. Markham, A. Birdwell, M. Arens, A. Saah, and L. Ratner.** 1992. Relationship to human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc. Natl. Acad. Sci. USA **89**:10247–10251.
43. **Meyerhans, A., R. Cheynier, J. Albert, M. Seth, J. Kwok, J. Sninsky, L. Morfeldt-Manson, B. Asjö, and S. Wain-Hobson.** 1989. Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. Cell **58**:901–910.
44. **Miller, A. D., and C. Buttimore.** 1986. Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. Mol. Cell. Biol. **6**:2895–2902.
45. **Miller, A. D., and G. J. Rosman.** 1989. Improved retroviral vectors for gene transfer and expression. BioTechniques **7**:980–990.
46. **Munis, J. R., R. S. Kornbluth, J. C. Guatelli, and D. D. Richman.** 1992. Ordered appearance of human immunodeficiency virus type 1 nucleic acids following high multiplicity infection of macrophages. J. Gen. Virol. **73**:1899–1906.
47. **Myers, G., B. Korber, J. A. Berzofsky, R. F. Smith, and G. N. Pavlakis.** 1991. Human retroviruses and AIDS. Los Alamos National Laboratory, Los Alamos, N.M.
48. **Niederman, T. M. J., J. V. Garcia, W. R. Hastings, S. Luria, and L. Ratner.** 1992. Human immunodeficiency virus type 1 Nef protein inhibits NF-κB induction in human T cells. J. Virol. **66**:6213–6219.
49. **Ratner, L., A. Fisher, L. L. Jagodzinski, H. Mitsuya, R. C. Gallo, and F. Wong-Staal.** 1987. Complete nucleotide sequences of functional clones of the AIDS virus. AIDS Res. Hum. Retroviruses **3**:57–69.
50. **Ratner, L., B. Starcich, S. F. Josephs, B. H. Hahn, E. P. Reddy, K. J. Livak, S. R. Petteway, Jr., M. L. Pearson, W. A. Haseltine, S. K. Arya, and F. Wong-Staal.** 1985. Polymorphism of the 3' open reading frame of the virus associated with the acquired immune deficiency syndrome, human T-lymphotropic type III. Nucleic Acids Res. **13**:8219–8229.
51. **Robert-Guroff, M., M. Popovic, S. Gartner, P. Markham, R. C. Gallo, and M. S. Reitz.** 1990. Structure and expression of *tat*-, *rev*-, and *nef*-specific transcripts of human immunodeficiency virus type 1 in infected lymphocytes and macrophages. J. Virol. **64**:3391–3398.
52. **Saag, M. S., B. H. Hahn, J. Gibbons, Y. Li, E. S. Parks, W. F. Parks, and G. M. Shaw.** 1988. Extensive variation of human immunodeficiency virus type-1 *in vivo*. Nature (London) **334**:440–444.
53. **Samuel, K. P., D. R. Hodge, Y.-M. A. Chen, and T. S. Papas.** 1991. Nef proteins of the human immunodeficiency viruses (HIV-1 and HIV-2) and simian immunodeficiency virus (SIV) are structurally similar to leucine zipper transcriptional activation factors. AIDS Res. Hum. Retroviruses **7**:697–706.
54. **Samuel, K. P., A. Seth, A. Konopka, J. A. Lutenberger, and T. S. Papas.** 1987. The 3'-orf protein of human immunodeficiency virus shows structural homology with the phosphorylation domain of human interleukin-2 receptor and the ATP-binding site of the protein kinase family. FEBS Lett. **218**:81–86.
55. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
56. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**:543–548.
57. **Simmonds, P., P. Balfe, J. F. Peutherer, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown.** 1990. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. J. Virol. **64**:864–872.
58. **Smith, M. S., K. Koerber, and J. S. Pagano.** Submitted for publication.
59. **Staden, R.** 1988. Methods to define and locate patterns of motifs in sequences. Comput. Appl. Biosci. **4**:53–60.
60. **Terwilliger, E., J. G. Sodroski, C. A. Rosen, and W. A. Haseltine.** 1986. Effects of mutations within the 3' *orf* open reading frame region of human T-cell lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenicity. J. Virol. **60**:754–760.
61. **Tsunetsugu-Yokota, Y., S. Matsuda, M. Maekawa, T. Saito, T. Takemori, and Y. Takebe.** 1992. Constitutive expression of the *nef* gene suppresses human immunodeficiency virus type 1 (HIV-1) replication in monocytic cell lines. Virology **191**:960–963.
62. **Varmus, H., and R. Swanstrom.** 1985. Replication of retroviruses, p. 75–134. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
63. **Vartanian, J.-P., A. Meyerhans, B. Asjö, and S. Wain-Hobson.** 1991. Selection, recombination, and G→A hypermutation of human immunodeficiency virus type 1 genomes. J. Virol. **65**:1779–1788.
64. **Vega, M. A., R. Guigo, and T. F. Smith.** 1990. Autoimmune response in AIDS. Nature (London) **345**:26.
65. **Werner, T., S. Ferroni, T. Saermark, R. Brack-Werner, R. B. Banati, R. Mager, L. Steinaa, G. W. Kreutzberg, and V. Erfle.** 1991. HIV-1 Nef protein exhibits structural and functional similarity to scorpion peptides interacting with K^+ channels. AIDS **5**:1301–1308.
66. **Yu, G., and R. L. Felsted.** 1991. Effect of myristoylation on p27 Nef subcellular distribution and suppression of HIV LTR transcription. Virology **187**:46–55.
67. **Zweig, M., K. P. Samuel, S. D. Showalter, S. V. Baden, G. C. DuBois, J. A. Lautenberger, D. R. Hodge, and T. S. Papas.** 1990. Heterogeneity of Nef proteins in cells infected with human immunodeficiency virus type 1. Virology **179**:504–507.