Molecular characterization of five human anti-human immunodeficiency virus type 1 antibody heavy chains reveals extensive somatic mutation typical of an antigen-driven immune response

(human antibodies/variable region gene segments/B-cell repertoire)

JENNIFER S. ANDRIS*, SYD JOHNSON†, SUSAN ZOLLA-PAZNER§, AND J. DONALD CAPRA¶

Departments of *Microbiology and 1Internal Medicine and 1Graduate Program in Immunology, The University of Texas Southwestern Medical Center, Dallas, TX 75235; †MedImmune, Inc., Gaithersburg, MD 20878; and §New York Veterans Affairs Medical Center and the Department of Pathology, New York University Medical Center, New York, NY 10016

Communicated by Richard M. Krause, May 6, 1991

ABSTRACT We report the heavy chain variable region sequences from the cDNAs of five previously described monoclonal cell lines producing human antibodies specific for the human immunodeficiency virus type 1 and detail the molecular characteristics, germ-line origins, and extent of somatic mutation among these antibodies. Three of the five heavy chain variable regions derive from the VHIV gene family, but each has arisen from a different heavy chain variable region (VH) gene segment within the VHIV family. In addition, one is derived from a VHII gene segment, and one is derived from a VHIV gene segment. Since four of the five antibodies arise from known germ-line VH elements, a precise determination of the extent of somatic variation is possible. The amount of variation from the closest germ-line sequence ranges from 4.5% to 14% and, among these antibodies, most of which is concentrated in the complementarity-determining regions. In general, the diversity (D) segments are long, characteristic of D-D fusions and/or extensive terminal deoxynucleotidyltransferase activity; however, definitive homologies cannot be found with the known germ-line D segments. Joining (JH) gene segment utilization appears random. The use of five different germ-line VH gene segments and extensive somatic mutation provides evidence that a polyclonal, antigen-driven immune response occurs during the natural infection with human immunodeficiency virus.

Immunoglobulin heavy and light chain variable regions are formed through the rearrangement of germ-line variable (VH), diversity (D), and joining (JH) gene segments and variable (VJ) and joining (JL) segments, respectively (1). The subsequent association of the two variable regions and the operation of several mechanisms of somatic variation provide the antibody specificity that is crucial in a humoral immune response.

In humans, the immunoglobulin heavy chain complex consists of variable region (VH) gene segments that have been divided into six families (2) (family members are >80% homologous to each other at the nucleotide level), >20 D segments (3-6), and six functional JH gene segments (4). The VH1, VHIII, and VHIIII families were originally designated as such from the amino acid sequences of myeloma proteins (7) and were later confirmed at the nucleotide level (5, 8-18). The smaller, more recently described VHIV, VHV, and VHIId rearrangements from chronic lympho-
efficiency virus type 1 (HIV-1) are reported here.** The cell lines were derived by EBV transformation of peripheral blood lymphocytes from HIV-positive individuals (34, 35). Three of the five variable regions are members of the V1H family, one derives from the V1H gene family, and one derives from the V2H gene family. By determining the nucleotide sequences to these antibodies, we hoped to assess the germ-line origin as well as the role of somatic variation present in antibodies that arise during the course of a natural infection that results from exposure to a relatively new human pathogen.

MATERIALS AND METHODS

Generation and Characterization of Human mAbs to HIV. Cell lines producing human mAbs to HIV were derived from peripheral blood cells from HIV-infected patients who were asymptomatic at the time the cell lines were developed. As previously described, lymphoblastoid cell lines were obtained by EBV transformation (34) or by fusion of EBV-transformed cells with the SHM-D33 heteromyeloma (35), generously provided by N. Teng. Cells producing antibodies specific for HIV were identified by reactivity on ELISA either with whole HIV lysates or with a 23-mer peptide whose sequence is identical to that which spans the tip of the V3 loop of HIVM1. Five mAbs from three patients were studied; these include 98-6 (IgG2) and 120-16 (IgG2), both of which are specific for a region in gp41 that maps to amino acids 644–663 (S.Z.-P., unpublished observations), 71-31 (IgG1), which is specific for p24, and 257-D (IgG1) and 268-D (IgG1), both of which are specific for the principal neutralizing domain of gp120 of MN and map to residues KRIHI and HIGPG, respectively. mAbs 257-D and 268-D have been shown to neutralize HIVMN but not HIVMN (35).

Oligonucleotides, First-Strand cDNA Synthesis, and PCR. Oligonucleotide primers and conditions used in the PCR amplifications and cDNA synthesis have been described (36).

Isolation, Cloning, and Sequencing of the Amplified Products. Amplified DNA was size selected on a 1% agarose gel, ligated into the Smal I site of pTZ19U phagemid vector, and transformed into CaCl2 competent BStJ22 bacteria. Single-stranded DNA used for sequencing was isolated from positive clones following supernatation with M13K07 as described (37). Sequencing was carried out on a minimum of four clones (both strands) using the dideoxy chain-termination method of Sanger (38) with a modified version of T7 DNA polymerase (Sequenase, United States Biochemical) (39). Clones 268-D and 257-D were also isolated by cDNA cloning and sequencing.

RESULTS

Cloning and Sequencing the Expressed VH Genes Encoding Anti-HIV-1 Antibodies. The complete variable region sequence of each antibody is shown in Fig. 1. The VH120-16 gene is a member of the VH2 J4 gene family and is most homologous to the previously described V71-2 germ-line gene (87.2%) (19). The VH98-6 gene segment is 85.2% identical to the V2-1 germ-line gene (19), another VH4 family member. The VH268-D gene is also a member of the VH2 J4 gene family and is 88.5% identical to the V71-4 germ-line gene (19). The VH region of antibody 71-31 derives from a VH1 germ-line gene, but the closest homology with any known VH1 gene (HG3) (8) is based on a 3-base-pair gap in CDR2. The last variable region, VH257-D, derives from a VH2 gene segment, VH21-2 (23) (95.5% homology).

**The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M67500 (120-16), M67501 (98-6), M67502 (71-31), M67503 (268-D), and M67504 (257-D)).

Fig. 2 shows each of the expressed variable regions compared to their likely, respective, germ-line equivalents. The percent nucleotide difference ranges from 4.5% to 14.8%. As expected for an antigen-driven immune response, there are numerous nucleotide differences within the CDRs. With the exception of VH268-D, there are an unusual number of nucleotide differences in the framework regions, in particular in framework III, when compared to the analogous germ-line sequences. This degree of variation exceeds the usual mutation rate for frameworks (i.e., 2%), suggesting a role for framework residues in antigen binding (40). More than 50% of the total mutations result in amino acid interchanges, suggesting an antigen-driven process.

Analysis of the D and JH Gene Segments. There are >20 known germ-line D segments and six functional JH segments in the human heavy chain complex. In expressed variable regions, the D segment (and the D-JH junction) comprises the third CDR and is of major importance in generating antibody specificity and diversity. We found little homology between the expressed D segments and the known germ-line D segments, suggesting, as others have postulated (32), that there are additional, as yet unknown, D segments. It is, therefore, difficult to distinguish between the actual D gene and those nucleotides that may represent N segment addition by terminal deoxynucleotidyltransferase (41). In the murine system, it has been shown that additional D segment diversity can be generated by inverted D segments, D-D fusions, and inverted D-D fusions (42). The 71-31 D segment represents a possible inverted D segment, as it is most homologous to a germ-line D segment, DM2 (6), in the opposite transcriptional orientation. Additionally, the D segments in antibodies 268-D and 257-D appear to result from the fusion of an inverted D segment to a second D segment. Another type of junctional diversity, termed "P" (palindromic) nucleotides, was first described in murine T-cell receptor genes to account for recurring mono- and dinucleotides at the V-J junction (43). There is one example of possible P nucleotide addition at the VH-D junction in antibody 120-16 where the presence of the dinucleotide, TC, creates a palindrome (Fig. 1A).

Among the five antibodies, three express the JH4 gene segment, one expresses JH2, and one expresses JH3. It is thought that the somatic "mutase" is random and can target sequences in the 5' and 3' untranslated regions as well as in the coding region of VH (12) and JH (44); therefore it is not unexpected that some nucleotides are different in the expressed JH segments when compared to their respective germ-line equivalents. Some of the germ-line JH segments have been shown to be polymorphic (36, 45). In all of the expressed JH4 gene segments reported here, an example of this polymorphism is seen where there is a "G" rather than an "A" at nucleotide 21 (Fig. 1B), nucleotide 9 (Fig. 1C), and nucleotide 21 (Fig. 1E) of the respective JH segments.

DISCUSSION

To date, the experimental data regarding the human immunoglobulin heavy chain repertoire come mainly from two sources: (i) the amino acid and nucleotide sequences of myeloma proteins and fetal and chronic lymphocytic leukemia VµDµJµ rearrangements, most of which are of unknown specificity (representing the bulk of the data), and (ii) monoclonal and polyclonal autoantibodies of known specificity. There is little structural information regarding antibodies to exogenous antigens (46–49). Therefore it is becoming increasingly important to further analyze the human B-cell repertoire to gain insight into mechanisms that regulate and determine VH (and VL) gene expression in different B-cell populations, particularly in immune responses to infection and/or vaccination. We report here the complete nucleotide sequences of the heavy chain variable regions of five human mAbs specific for HIV-1. Of these five
antibodies, three belong to the V_{H} IV gene family, one derives from a V_{H}I gene segment, and one derives from a V_{H} V gene. Comparison of these variable region sequences with their likely germ-line equivalents suggests that they arose from separate clones and resulted from a potent antigen-driven process.

To understand the process by which an expressed variable region has evolved, it is critical to know its germ-line origin.
Immunology: Andris et al.


FIG. 2. Nucleotide sequence of each expressed variable region compared to the nucleotide sequence of the closest germ-line genes. CDR1 and CDR2 are indicated by the asterisks. (A) Nucleotide sequence of VH120-16 compared to the VH12 (VH1V) (19) germ-line sequence. There are 46 nucleotide differences between the two sequences. (B) Nucleotide sequence of VH98-6 compared to the VH2-1 (VHIV) (19) germ-line sequence. There are 55 nucleotide differences between the two sequences. (C) Nucleotide sequence of VH71-31 compared to the HG3 (VHII) (8) germ-line sequence. There are 27 nucleotide differences between the two sequences and a 3-base-pair gap in CDR2. (D) Nucleotide sequence of VH268-D compared to the VH7-4 (VHIV) (19) germ-line sequence. There are 42 nucleotide differences between the two sequences. (E) Nucleotide sequence of VH257-D compared to the VH251 (VHII) (23) germ-line gene. There are 17 nucleotide differences between the two sequences.

In most cases, assigning a germ-line counterpart to an expressed gene is difficult due to the large number of gene segments in the VHI and VHIll families, many of which are unknown at the germ-line level. With the discovery of the VH1V, VH1V, and VHVI gene families, assigning a germ-line gene to an expressed sequence is less difficult because the number of germ-line gene segments in these families seems to be saturated. Since four of the five antibodies studied here derive their VH segments from the VH1V and VH1V gene families, we are also better able to address the extent of somatic variation. When each expressed sequence is compared to the germ-line gene with the closest homology, the sequences reveal extensive, apparent somatic mutation, most of which is concentrated in the CDRs. In most of these antibodies, there is also an unusual number of nucleotide differences within the framework regions as well, indicating a possible role for some framework amino acids in antigen binding.

The molecular characterization of these HIV antibodies not only provides insight into the "normal" adult B-cell repertoire but also presents an opportunity to examine the nature of an immune response in hyperimmunized individuals. A chronic infection, such as HIV, provides a situation in which there is continual and vigorous antigenic stimulation, as evidenced by the fact that antibody-producing cells circulate in the peripheral blood (50) and antibody titers to viral components in HIV-infected individuals are extraordinarily high (51). Under such circumstances, one might expect to find antibodies resulting from antigen-driven selection over a period of months or even years. The data presented here support this view. Although we have not completely ruled out the existence of additional germ-line VHI gene segments, the extent of variation is striking in the CDRs and framework regions. Generally somatic mutation is thought to occur randomly in immunoglobulin variable regions, but it is largely seen in the CDRs due to antigen selection. Although the five variable regions presented here are also mutated in the framework regions, in no instance are the frameworks more mutated than the CDRs. Fig. 3 shows the deduced amino acid sequence of 257-D compared to the deduced amino acid sequence of the VH1V germ-line gene, VH251 (23). There are only three VH1V germ-line genes and it is clear that VH251 has given rise to this expressed antibody. There is a total of five nucleotide differences (1.7%) (Fig. 2E) within the framework (all of which result in a change in the amino acid),
which is close to the expected framework variation of 2.0%, whereas there is 16–22% variation in the two CDRs—10 times the framework variation. This antibody provides a striking example of selection by antigen.

Early studies demonstrated that HIV immune sera from infected patients did not prevent HIV infection in chimpanzees that were challenged with virus (52). In contrast, more recent studies have demonstrated that HIV hyperimmune human immunoglobulin preparations can passively protect chimpanzees that are challenged with human T-lymphotrophic virus type III B (53). Similarly, polyclonal chimpanzee IgG with HIV neutralizing activity has been shown to protect against challenge with the virus (54). Since some antibodies that bind HIV have been shown to enhance the infection, rather than neutralize the virus (55), these conflicting results may be explained, in part, by the heterogeneous nature of the immunoglobulin preparations used—i.e., protection may depend on the presence of enhancing antibodies. The molecular characterization of the HIV-1 antibodies studied here provides an opportunity for the expression and production of a homogeneous preparation of neutralizing antibodies.

Finally, these data, in addition to other, unpublished observations, indicate that a majority of the human B-cell repertoire to exogenous antigens arises from a distinct set of germ-line genes rather than from the genes used by the fetal, leukemic, and autoimmune repertoires (33). This suggests that at least some V_{H} germ-line genes are reserved for the latter purposes and may play a regulatory role in the expression of other V_{H} genes used in a normal immune response. If such a V_{H} gene utilization bias exists it would indicate that repertoire development is not as random as initially believed. The analysis of additional antibodies specific for exogenous antigens should provide further insight into this matter.

We thank Dr. Virginia Pascual for many helpful discussions regarding the work and David Cohen for his work on the 98-6 cell line. J.D.C. holds the Edwin L. Cox Distinguished Chair in Immunology and Genetics at the University of Texas Southwestern Medical Center and is also supported in part by grants from the National Institutes of Health (AI12127, CA44016, GM31689) and the Welch Foundation. S.Z.-P. is supported in part by funds from the National Institutes of Health (AI127638) and the Department of Veterans Affairs. J.S.A. is supported in part by the Texas Department Ladies Auxiliary, Veterans of Foreign Wars.