ABSTRACT The ability to respond to specific antigens develops in a programmed fashion. Although the antibody repertoire in adults is presumably generated by stochastic combinatorial joining of rearranged heavy variable, diversity, and joining (VH-DH-JH) and light (VL-JL) chains, experimental evidence in the mouse has shown random nonelection of variable gene segments during ontogeny and in response to specific antigens. In this study, we have performed sequence analysis of 104-day human fetal liver-derived, randomly isolated constant region C2 transcripts and demonstrate a consistent preference during fetal life for a small subset of the highly conserved Vp3 family gene segments. In addition, the data show that this preferential gene segment utilization extends to the DpQ52 and Jp3 and Jp4 loci. Sequence analysis of two “sterile” DH-JH transcripts suggests that transcriptional activation of the Jp-proximal DpQ52 element may precede initiation of DH-JH rearrangement and influence fetal Dp utilization. Sequence comparisons reveal striking nucleotide polymorphism in alllic gene segments which is poorly reflected in the peptide sequence, implying considerable evolutionary selection pressure. Although vertebrate species utilize a variety of strategies to generate their antibody repertoire, preferential utilization of Vp3 segments is consistently found during early development. These data support the hypothesis that Vp3 gene segments play an essential role in the development of the immune response.

Immunoglobulins are generated by combinatorial joining of rearranged gene segments of the heavy chain variable, diversity, and joining regions (VH, DH, and JH) and light chain regions (VL and JL) (1). Starting with less than 1000 of these germ-line elements, more than 109 different antigen binding sites can be generated even in the absence of either junctional diversity or somatic mutation. In mice, only a small fraction of this potential repertoire seems to be expressed as functional antibody (2). Utilization of this repertoire appears to be developmentally regulated in a strain-specific fashion (3–5) with subsequent modification by environmental stimuli (7).

The human newborn is relatively immunodeficient at birth (8). Controlled mobilization of germ-line variable gene segments has been postulated to underlie, in part, the maturation of humoral immunity (5, 9). With these observations in mind, we have concentrated on the use of molecular cloning strategies to dissect the development of the heavy chain repertoire during fetal life. Because the extent of VH region polymorphism in the outbred human population is undefined, we have chosen to analyze individual fetal samples.

Fetal B lymphopoiesis begins in the liver, with pre-B cells first detectable by 8 weeks of gestation (10). We isolated fetal liver mononuclear cells, which are rich in pre-B cell precursors, and generated oligo(dT)-primed cDNA libraries. We previously reported evidence of preferential usage of VH, DH, and JH gene segments by cloning and sequencing 15 JH-containing constant region C2 heavy chain transcripts from a 130-day-gestation cDNA library (11). The 14 VH-containing cDNAs represented only 9 germ-line gene segments, 5 of which belonged to the Vp3 family. Preference was shown for 1 (DpQ52) of more than 20 germ-line Dh (12, 13) and 2 (Jp3 and Jp4) of 6 functional germ-line Jh elements (14). In this communication, we extend our analysis of heavy chain transcripts present in the 130-day library and also analyze a 104-day library (9) for comparison with our previous results and to determine the extent of restriction during an earlier stage in fetal life.

Striking similarities in the heavy-chain repertoire expressed by two unrelated human fetuses of 4–5 months gestation indicate the existence of a conserved B-cell developmental program of heavy chain variable element expression.

MATERIALS AND METHODS

Human fetal liver samples were obtained from a karyotypically normal anencephalic abortus at 130 days of gestation, and from a second 104-day abortus with a neural tube abnormality (both gifts of T. Shepard, University of Washington, Seattle). The isolation of mononuclear cells from these tissue samples, purification of poly(A)+ RNA, generation of oligo(dT)-primed cDNA libraries, and sequencing of C2+ cDNAs have been previously described (9, 11).

RESULTS

Nineteen C2+ clones were detected in a total of 8.5 x 105 recombinants. One-third of the clones contained identical 5′ nonvariable sequences with numerous stop codons in all three reading frames. The presence of these “sterile” sequences is common in early lymphoid cells (15). In two clones, reverse transcription terminated in the middle of the Jh gene segment. Another contained an unusual nontranslated sterile sequence which will be reported elsewhere. The unique sequence at the site of VH->DH->JH joining in each of the remaining 10 clones demonstrates that each transcript was derived from an independent gene rearrangement event.

Eight of the 10 VH->DH->JH clones include complete VH coding sequences (Fig. 1). Of the two incomplete clones, one (M44) ends within the second hypervariable region (CDR II), and the other (M61) contains all but the most amino-terminal portion of framework I. In mouse lymphocytes nonfunctional

Abbreviations: V, variable; D, diversity; J, joining; C, constant; VH, heavy; VL, light; CDR, complementarity-determining region.
*To whom reprint requests should be addressed.
†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M34020 for clone 70p1, M34021 for clone 74p1, M34022 for clone 83p2, M34023 for clone M26, M34024 for clone M43, M34025 for clone M44, M34026 for clone M49, M34027 for clone M60, M34028 for clone M61, M34029 for clone M71, M34030 for clone M72, M34031 for clone M74, and M34032 for clone M85.
transcripts have much lower steady-state levels than functional immunoglobulin mRNAs (17, 18), presumably reflecting reduced mRNA stability. This bias is apparent in our sample of 10 transcripts, which contains only 1 nonproductive \( \text{VH}_4-\text{DH}-\text{JH} \) join. Thus, as in our 130-day library (11), the sequences reported here sample primarily the translatable repertoire of heavy chains.

Human \( \text{VH} \) sequences can be grouped into six families on the basis of \( \approx 80\% \) shared nucleotide identity (19, 20). With the exception of two previously unreported \( \text{VH}2 \) gene segments (transcripts M60 and M44) and a \( \text{VH}5 \) transcript (M61), the \( \text{VH} \) repertoire of this unrelated 104-day fetus overlaps the 130-day \( \text{VH} \) repertoire (Fig. 2). The single member \( \text{JH} \)-proximal \( \text{VH}6 \) family is used in both libraries (clones 15pl (130-day) and M71 (104-day)). However, 60% of the 104-day \( \text{VH} \) transcripts (Fig. 1) belong to the human \( \text{VH}3 \) family, which was also preferred at 130 days. Two of these \( \text{VH}3 \) transcripts (clones M72 and M85) differ from both of the previously reported 130-day sequences and companion 104-day sequences by one and three bases, respectively (Fig. 1A).

Note that the single base pair change in clone M72 does not result in a peptide substitution, nor do two of the three base changes in clone M85 (Fig. 1B). These two \( \text{VH}3 \) gene segments presumably represent alleles of the germ-line \( \text{VH} \) gene segments first identified by their presence in the 130-day transcripts 5pl and 20pl (Fig. 2) (11). The \( \text{VH}3 \) transcript M49 accesses a nearly identical \( \text{VH}3 \) pool, even though it cannot form a functional peptide product. Therefore, mechanisms which result in preferential use of these \( \text{VH}3 \) elements must act, at least in part, at the nucleic acid level. These findings confirm the existence of a consistent preference for specific \( \text{VH} \) elements in the early human antibody repertoire.

In the 10 104-day \( \text{VH}2-\text{DH}1-\text{JH} \) transcripts, there is favored use of \( \text{JH}3 \) (4 sequences) and \( \text{JH}4 \) (3 elements), although \( \text{JH}2, \text{JH}5, \), and \( \text{JH}6 \) elements are each used once (Figs. 3 and 4). Relative frequencies of \( \text{JH} \) usage in the adult are unknown,

---

**Fig. 1.** (A) Nucleotide sequences of 10 \( \text{VH} \) regions from a fetus of 104 days of gestation (designated by the prefix "M") and 5 human fetal \( \text{VH} \) regions from a fetus of 130 days of gestation (15pl, 20pl, 30pl, 5pl, and 83pl) are compared to clone 30pl, with a dot denoting nucleotide identity. Sequences begin at codon 1 of the processed \( \text{VH} \) region. Complete cDNA sequences are available through GenBank. The 3' terminus of each \( \text{VH} \) is arbitrarily defined as codon 93 (16). Sequences are grouped by family; and, within family, by \( \text{VH} \) gene segment identity. Presumed allelic differences in clones M72 and M85 are underlined. \( \text{VH} \) family assignment is on the left. (B) Amino acid sequence homologies among the 13 human fetal \( \text{VH} \) clones. The translation products of each cDNA clone are presented in single-letter code, aligned with clone 30pl as in A. The framework generating the nontranslatable rearrangement seen in clone M49 is marked by a --. The single polymorphic amino acid residue in the complementarity-determining region (CDR) 1 of clone M85 is underlined.

---

**Fig. 2.** Twenty-four \( C_\gamma \) independent \( \text{VH}2-\text{DH1}-\text{JH} \) transcripts have been randomly cloned and sequenced from two unrelated fetuses of 104 and 130 days of gestation. The 12 \( \text{VH} \) elements which contributed to these rearrangements are grouped by family and identified by a representative cDNA clone.
but in the fetus utilization clearly appears nonrandom.

The third hypervariable region (CDR III) is generated by V₃₉₄-D₃₉₄-H₃₉₄ joining. Random nucleotide addition (N regions) and junctional flexibility makes assignment of D₃₉₄ origin problematic. For example, four clones (M44, M49, M61, and M85) are unassignable (Fig. 4), and transcript M43 shares five bases with D₃₉₄QS2, seven bases with D₂₂/12 (12), and eight bases with DM1 (13). Two CDR III regions appear to contain members of the DN1 family (13): clone M72 shares 14 bases of identity with DN1, and clone (M26) shares 19 bases of interrupted identity. In our 130-day library, 8 of 14 CDR III regions shared between five and nine bases of identity with D₃₉₄QS2 (14). In the current sample of 10 clones, 3 (M60, M71, and M74) share between 8 and 10 base-pair identity with D₃₉₄QS2. The probability that these three transcripts would be drawn from the D₃₉₄QS2 locus at random is <3 × 10⁻⁵.

To extend our D₃₉₄ locus analysis, we examined 13 C₄ᵦ clones which had been randomly isolated from an additional 3.5 × 10⁵ 130-day library recombinants but had not been evaluated (11). Six of the 13 clones did not hybridize to either our sterile specific oligonucleotide [5'-GCCCAAGACTGT-CATGGGTATACA-3' (11)] or our cocktail of three V₃₉₄ oligonucleotides [5'-GCACAAGTATACAG-GCCGTGTC-3', 5'-ACAGTAATACGCCGCTGTC-3', and 5'-TGACAGTATACAGCCTGGTTC-3'] (21) which identify >95% of known germ-line V₃₉₄ gene segments (data not shown). Four of these clones contained further D₃₉₄-H₃₉₄ information. The first (70p1) utilized J₃₉₄6 and terminated at the site of D₃₉₄-J₃₉₄ joining. The second (83p2) contained a V₃₉₄5 gene segment identical to the V₃₉₄5 element in the 104-day clone M61 (Fig. 1), as well as to the germ-line V₃₉₄ element 5-1R1 (19).

Note that the CDR III region of this V₃₉₄5 transcript shares 8 bases of identity with D₃₉₄QS2 (Fig. 4). The third clone (74p1) was an authentic D₃₉₄ transcript joining D₃₉₄QS2 to J₃₉₄2 with an intervening N region sequence (Fig. 5). The fourth clone (84p1) represents a transcript which initiates 5' to the J₃₉₄ region; reads through the unarranged pseudo-J₃₉₄1, D₃₉₄QS2, and J₃₉₄1 gene segments; and then splices appreciably to the 5' terminus of C₃₉₄ (Fig. 5). These transcripts provide further evidence for preferential utilization of D₃₉₄QS2.

DISCUSSION

Fetal D₃₉₄ and J₃₉₄ Elements Are Preferentially Rearranged.

Our data indicate a strong preference for D₃₉₄QS2 and J₃₉₄3 and J₃₉₄4 utilization in the second trimester antibody repertoire. Further support for nonrandom D₃₉₄-J₃₉₄ rearrangement can be found in analysis of D-J joins in human fetal B-lineage cells at 19 weeks gestation transformed by Epstein-Barr virus (22). A striking gradient of D₃₉₄QS2 rearrangement to J₃₉₄1 > J₃₉₄2, > J₃₉₄3, > J₃₉₄4, 6 is seen. J₃₉₄ representation in our two D₃₉₄-H₃₉₄ transcripts also reflects D₃₉₄QS2 proximity (Fig. 5).

![Fig. 3. J₃₉₄ utilization in 25 C₄ᵦ clones randomly isolated from 104- and 130-day fetal liver-derived monoclonal cell cDNA libraries. A consistent preference for J₃₉₄3 and J₃₉₄4 was detected.](image)

![Fig. 4. D₃₉₄ and J₃₉₄ sequences from human fetal C₄ᵦ cDNA clones. (A) Nucleotide sequences from the D₃₉₄/N region of 13 fetal heavy chain transcripts. Sequences are aligned with their germ-line counterparts where identification is possible: D₃₉₄QS2 (14), D₃₉₄1 (13), D₃₉₄1, and D₂₂/12 (12). A dot denotes sequence identity with the given germ-line segment. Clone M43 is shown three times, demonstrating the difficulty of D₃₉₄ assignment in a completely rearranged variable element. The heptamer recombination region is underlined in the D₃₉₄-J₃₉₄ joining transcript 70p1, as well as in the incomplete J₃₉₄-containing transcript 70p1 from the 104-day library (11). (B) Nucleotide sequences from the J₃₉₄ region indicating the rich polymorphism present in the human population. Sequences are grouped by J₃₉₄ identity. Dots denote homology to J₃₉₄1 (14). Single base substitutions from previously published sequences (14) are underlined. Clones 13P1, 51P1, 56P1, and 2P1 have been previously reported (11) and are included to illustrate the range of polymorphism seen in this sampling of only five human haplotypes.

The accessibility model propounded by Alt and co-workers (23) posits that a gene can undergo rearrangement only when it is physically accessible to the recombinase. Gene segment accessibility is associated with, and may in fact require, prior transcription of the rearranging element (23). Clone 84p1 (Fig. 5) contains a novel sterile transcript initiating upstream of D₃₉₄QS2, reading through the recombination signals to J₃₉₄1, and then splicing to C₃₉₄. In keeping with the accessibility model, this transcript may precede initial D₃₉₄-J₃₉₄ rearrangement in the pro-pre-B cell and thus could represent the earliest activation event preceding heavy chain rearrangement. Note that few of our transcripts contain J₃₉₄1, whereas V₃₉₄-containing transcripts (e.g., clones 83p2, M74, and M71, Fig. 4) utilize D₃₉₄QS2 in association with J₃₉₄2, J₃₉₄3, and J₃₉₄4, respectively. The relative paucity of J₃₉₄1 and J₃₉₄2 utilization in the expressed V-D-J rearrangements suggests either that additional rounds of D₃₉₄-J₃₉₄ rearrangement may precede V₃₉₄ splicing or that access to additional D₃₉₄ and J₃₉₄ gene segments may parallel access to the rest of the V₃₉₄ locus.
Even though use of VH, Dm, and JH elements is selective, junctional flexibility ensures a diverse fetal repertoire. For example, DmQ52 is read in all three frameworks and, in concert with N region addition, generates three completely different CDR III regions in M60, M71, and 83P2. This dependence on the third hypervariable region is reminiscent of diversity generation in T-cell receptor rearrangements (24).

**Human Heavy Chain Gene Segments Demonstrate Excessive Nucleotide Polymorphism yet Preserve Peptide Sequence.** The JH gene segment codes for the conserved framework IV region (codons 101–113) and can contribute up to nine codons of the highly variable CDR III region (16). Combining the original JH region sequence of Ravetch et al. (14) with our two samples, five haplotypes have been sampled. Within this limited set of observations, four alleles can be distinguished for JH3, two for JH4, three for JH5, and three for JH6 (Fig. 4). Six of the eight base pair differences are located in framework IV and are silent. Conversely, the three base pair differences in the CDR III region can both result in amino acid substitutions. Note that identification of allelic JH segments demonstrates similar heavy-chain preferences in both chromosomes. Similarly, only one of the four nucleotide differences in the two VH alleles results in an amino acid change. Hybridization analysis with V-region specific oligonucleotides has also suggested extensive heterohaplony and polymorphism in the human heavy chain locus (25). These data imply that the observed polymorphism may primarily reflect single base pair changes and that the conserved peptide sequences are under significant selection pressure.

**Fetal VH Utilization Is Consistently Restricted with Striking Preference for Specific VH3 Elements.** We have identified 24 individual VH-DH-JH transcripts randomly isolated from two independent cDNA libraries derived from early second-trimester fetal B-lineage cells; 70% of the sequences obtained from the 104-day library are also present in the 130-day set. The human germ-line VH repertoire may contain more than 100 segments per haploid genome (19, 20). Hybridization of VH probes to 8-week fetal liver-derived mononuclear cell RNA suggests that VH utilization is limited to VH3 and VH6 family gene segments (26) when B-lineage development is initiated. Northern analysis of Epstein–Barr virus-transformed cell lines derived from fetal liver and spleen suggested VH utilization might be representative of the germ-line repertoire by 19 weeks with 60% of the clones expressing VH3 elements (27). In our sample, however, although 60% of the fetal cDNAs contain VH3 elements, sequence analysis reveals that these transcripts are drawn from a subset of only 5 individual gene segments out of an estimated total pool of at least 25–30. Note that the VH1 family contains a minimum of 20–25 members: VH1-2, 5–10; VH1-4, 6–10; VH1-5, 2–3; and VH1-6, 1 member (reviewed in ref. 27). Thus, a small cohort of VH3 elements consistently contributes disproportionately to the fetal antibody repertoire.

The ability to generate antibody responses to specific vaccines follows a predictable course during human infancy. If developmentally controlled restriction of VH gene segment utilization regulates, in part, this observed hierarchy of antigen responsiveness, we predict that the expressed antibody repertoire during fetal life would be similar between fetuses and that limitations in the repertoire would continue into early neonatal life. The striking consistency which we have found in these two fetal samples supports the hypothesis that control of variable region utilization contributes to the relative immunodeficiency of the fetus and leaves open the question of whether or not these predictable limitations may extend into neonatal life.

**Preference for VH3-Like Gene Segments Is Also Seen in Fetal Mice.** VH families arose prior to the mammalian radiation and have since been conserved (28, 29). This conservation appears to reflect selection at the level of protein sequence, and the conserved regions are discretely localized on a solvent-exposed face of the heavy chain, at some distance from the classic antigen-binding site (28). Analysis of hybridomas derived from mouse fetal pre-B cells (5, 30) and Abelson virus-transformed pre-B cell lines have demonstrated a preference for members of the VH7183 family (30), corresponding to less than 10% of the total germ-line VH repertoire. It is remarkable that the most highly homologous VH3 gene segments in the mouse and human germ-line repertoires (28, 32) known to us [the VH7183 segment VH3E415 (30) and the VH3 segment 30p1 (11)] are both favored constituents of the early fetal repertoire. JH Proximity Contributes to Early VH Gene Segment Selection but Is Likely Not the Only Factor. In the mouse, fetal VH is rearranged to VH-JH rearrangements show marked correspondence to the chromosomal order of the gene segments involved (5, 30). In BALB/c, the VH7813 family is positioned at the proximal end of the VH locus (30), whereas members of the later-expressed J558 family are more distal. Within the 7183 family, the most JH proximal gene segment, VGH81X, is preferentially rearranged (30). These data form the basis of...
scanning or chromatin potentiation regulatory models postulating initiation of a V_{H} gene segments proximal to the D_{H}-V_{H} locus (5, 30) with subsequent linear access to the more J_{H}-distal V_{H} elements.

In humans, the two most J_{H}-proximal functional V_{H} elements are the unique V_{H}6 (21, 33) and a V_{H}3 (34) gene segment. These elements may represent the favored target for rearrangement at 8 weeks of gestation. However, although present, they are not the most commonly utilized elements in these two second-trimester samples. Provocatively, recent observations in the mouse also suggest that the favored V_{H}81X gene segment may not be the most J_{H}-proximal V_{H} (35). Although a common recombinase is postulated for all antigen receptor rearrangements, we have recently demonstrated striking evolutionary conservation of V_{H} family-specific sequence involving the two-turn DNA spacer region between the heptamer and nonamer recombination signals (28). These conserved intervening sequences could influence recombinase activity either directly through differential intermediate complex stability or through associated binding factors. Selective rearrangement has also been reported in some cultured cell lines (36). These observations suggest that control of this developmental program may involve multiple regulatory mechanisms.

Early Use of V_{H}3 Elements Is Common in All Vertebrates, Although the Mechanisms Which Influence the Choice Vary. Horned sharks contain multiple V_{H}-D_{H}-J_{H}-C_{H} clusters (37). In the light chain locus in birds, only a single functional V region is found (38) and diversity is generated by gene conversion from upstream pseudogenes. Both birds (39) and rabbits (40, 41) have multiple V_{H} gene segments, but all are V_{H}3-like. Therefore, in spite of the different mechanisms which are utilized to generate adult diversity in this broad range of species usage of V_{H}3-like elements appears to be a conserved component of the earliest antigen reponse. These observations imply that V_{H}3-like elements play an important biologic role. One possibility is that these sequences have been selected in response to antigens shared by common pathogens. However, the structural regions which exhibit greatest conservation do not coincide with the classic antigen-binding site (28). Studies in both mice and humans have demonstrated elevated autoactivity among neonatal antibody repositories. Two of the three V_{H}3-like elements we have studied are most commonly utilized (30p1, 20p1) can form self-reactive antibodies (42, 43). Similarly, early antibody repositories appear enriched for self-reactivity (6). Studies have shown that perturbation of B cells with anti-idiotypic antibodies during critical windows of development can alter the adult repertoire (6). These observations suggest that V_{H}3 elements play a key role in the establishment of the immune system in several classes of vertebrates. If so, alterations in the developmental program of variable element utilization may contribute to autoimmunity, immunodeficiency, or other immune dysfunction.

This work was supported in part by National Institutes of Health Grant AI23694. H.W.S. is an RJR Nabisco Research Scholar in Immunology.