Nucleic Acids Research, 1992, Vol. 20, No. 23 6287–6295

A transgenic mouse that expresses a diversity of human sequence heavy and light chain immunoglobulins

Lisa D.Taylor, Condie E.Carmack, Stephen R.Schramm, Roshanak Mashayekh, Kay M.Higgins, Chiung-Chi Kuo, Clive Woodhouse, Robert M.Kay and Nils Lonberg*
GenPharm International, 2375 Garcia Avenue, Mountain View, CA 94043, USA

Received August 27, 1992; Revised and Accepted October 26, 1992

ABSTRACT

We have generated transgenic mice that express a diverse repertoire of human sequence immunoglobulins. The expression of this repertoire is directed by light and heavy chain minilocus transgenes comprised of human protein coding sequences in an unrearranged, germ-line configuration. In this paper we describe the construction of these miniloci and the composition of the CDR3 repertoire generated by the transgenic mice. The largest transgene discussed is a heavy chain minilocus that includes human μ and $\gamma 1$ coding sequences together with their respective switch regions. It consists of a single 61 kb DNA fragment propagated in a bacterial plasmid vector. Both human heavy chain classes are expressed in animals that carry the transgene. In light chain transgenic animals the unrearranged minilocus sequences recombine to form VJ joints that use all five human J_κ segments, resulting in a diversity of human-like CDR3 regions. Similarly, in heavy chain transgenics the inserted sequences undergo VDJ joining complete with N region addition to generate a human-like V_H CDR3 repertoire. All six human J_H segments and at least eight of the ten transgene encoded human D segments are expressed. The transgenic animals described in this paper represent a potential source of human sequence antibodies for *in vivo* therapeutic applications.

INTRODUCTION

A number of investigators have previously reported immunoglobulin and T-cell receptor gene sequence rearrangements in transgenic animals. Bucchini *et al.* (1) reported the rearrangement of a germline chicken λ light chain locus inserted into transgenic mice. Similarly, Goodhart *et al.* (2) observed the rearrangement of a rabbit kappa light chain construct in mice. Two other groups (3,4) have made transgenic animals with chimeric rearrangement test constructs and shown that the transgenes rearrange during lymphoid cell development. Bruggemann *et al.* (5) generated mice containing a hybrid human/mouse heavy chain minilocus construct. This construct included one human and one mouse V segment, three mouse D segments (two of which had been altered by site directed mutagenesis to appear human), one human D segment, all six human J segments, and a chimeric human/mouse μ gene. The authors observed rearrangement of the transgene sequences in spleen and thymus as well as serum expression of human μ epitopes. No analysis of the structure of the rearrangements was reported. Bruggemann *et al.* subsequently reported the generation of transgenic mice by the co-injection of two cosmid clones (6). Together, these clones encompassed 100 kb of the human heavy chain locus and included most of the D region as well as the entire human J and μ regions. The constructs included 2 functional human V segments. The authors observed rearrangement of the transgene in lymphoid tissue; however, sequence analysis of the resulting VDJ joints showed only short CDR3 sequences with no recognizable human D segments.

In this report we describe transgenic animals that carry both light and heavy chain minilocus constructs comprised of human coding sequences. The heavy chain construct encodes two different isotypes, μ and $\gamma 1$. The transgenic mice that we have generated express a diversity of human sequence immunoglobulins that incorporate all of the human $J\kappa$ and J_H segments, at least eight different human D segments, and two different human heavy chain constant region segments. The human light and heavy chain CDR3 repertoires of these animals are comparable with authentic human CDR3 repertoires.

MATERIALS AND METHODS

Plasmid vectors

For the purpose of building very large transgene constructs in bacterial plasmids, we have developed a series of new cloning vectors. These vectors contain different polylinker sequences cloned into the NotI site of pGP1a, the first vector in the series. We generated pGP1a by ligating two synthetic oligonucleotides, caa gag ccc gcc taa tga gcg ggc ttt ttt ttg cat act gcg gcc gct and aat tag cgg ccg cag tat gca aaa aaa agc ccg ctc att agg cgg gct, into EcoRI/StyI digested pBR322. The resulting plasmid, pGP1a, is designed for cloning very large DNA constructs that can be excised by the rare cutting restriction enzyme NotI. It contains a NotI restriction site downstream (relative to the ampicillin resistance gene, AmpR) of a strong transcription termination

* To whom correspondence should be addressed

signal derived from the trpA gene (7). The vectors pGP1b, pGP1c, pGP1d, and pGP1f were derived from pGP1a and contain different polylinker cloning sites. The polylinker sequences are: pGP1a, GCG GCC GC; pGP1b, GCg gcc gcc tcg aga tca cta tcg att aat taa gga tcc agc agt aag ctt gcG GCC GC; pGP1c, GCg gcc gca tcc cgg gtc tcg agg tcg aca agc ttt cga gga tcc gcG GCC GC; pGP1d, GCg gcc gct gtc gac aag ctt atc gat gga tcc tcg agt gcG GCC GC; pGP1f, GCg gcc gct gtc gac aag ctt cga att cag atc gat gtg gta cct gga tcc tcg agt gcG GCC GC. The heavy chain minilocus constructs were built in a plasmid vector derived from pGP1b that also contains the rat immunoglobulin 3′ heavy chain enhancer (8). This enhancer was amplified from rat liver DNA using the following two synthetic oligonucleotides as primers: ctc cag gat cca gat atc agt acc tga aac agg gct tgc and ctc cag gat cca gat atc agt acc tga aac agg gct tgc. The amplified product was digested with BamHI and SphI and cloned into BamHI/SphI digested pNNO3 (R. Tizard, Biogen, Cambridge, MA), a pUC derived plasmid that contains a polylinker with the following restriction sites, listed in order: NotI, BamHI, NcoI, ClaI, EcoRV, XbaI, SacI, XhoI, SphI, PstI, BglII, EcoRI, SmaI, KpnI, HindIII, and NotI. The resulting plasmid, pRE3, was digested with BamHI and HindIII, and the insert containing the rat Ig heavy chain 3′ enhancer cloned into BamHI/HindIII digested pGP1b to generate pGPe.

Heavy chain miniloci

Isolation of human μ sequences. We screened a λ phage library of human genomic DNA sequences with human μ specific oligonucleotide probes and isolated clones that spanned the J-μ-δ region. We combined three different fragments to generate the plasmid pJM2: a 6 kb HindIII/KpnI fragment containing all six J segments as well as D segment DHQ52 and the heavy chain J-μ intronic enhancer, the adjacent downstream 10.5 kb HindIII/XhoI fragment, containing the μ switch region and all of the μ constant region exons, and a 4 kb XhoI fragment that contains sequences downstream and includes the so-called Σμ element involved in μ deletion in certain IgD expressing B cells (9, 10).

Isolation of human D region sequences. We used human D region specific oligonucleotides to isolate phage clones containing the D1 and D2 portions of the human D region. A 5.5 kb XhoI fragment, that includes the D elements D_{K1}, D_{N1}, D_{IR2}, D_{M2}, and D_{LR2} (11), was combined with the adjacent upstream 5.2 kb XhoI fragment that includes the D elements D_{LR1}, D_{XP1}, $D_{XP'1}$, and D_{A1}, to give the plasmid pDH1. pDH1 and pJM2 were combined to create the plasmid pCOR1. Plasmid pCOR1 was partially digested with XhoI and a 10.3 kb genomic HindIII fragment containing the functional human heavy chain variable region segments V_H251 and the variable segment pseudogene V_H105 (12) inserted upstream to produce the transgene construct pIGM1 (Figure 1). The plasmid pIGM1 contains a single functional human variable region segment, at least 10 human D segments, all 6 human J_H segments, the human J-μ enhancer, the human sm element, the human μ switch region, all of the human μ coding exons, and the human Sm element, together with the rat heavy chain 3′ enhancer.

Isolation of γ1 constant region sequences. We isolated human γ1 genomic clones from a phage library using specific oligonucleotide probes and confirmed by DNA sequence analysis that the clones belonged to the γ1 subclass. We combined three adjacent genomic fragments, a 5.3 kb HindIII fragment, a 7.6 kb HindIII/BamHI fragment, and a 4.5 kb BamHI fragment to generate the plasmid clone pγe2. pγe2 contains all of the γ1 constant region coding exons, and the upstream switch region and sterile transcript exons, together with 5 kb of downstream sequences, linked to the rat heavy chain 3′ enhancer. This clone contains a unique XhoI site at the 5′ end of the insert. The plasmid pIGM1 was digested with XhoI and the 43 kb insert isolated and cloned into XhoI digested pγe2 to generate the plasmid pHC1 (Figure 1).

Light chain minilocus

$V_κ$ gene. We screened a human genomic DNA phage library with the $V_κ$ light chain specific oligonucleotide probe 5′- agg ttc agt ggc agt ggg tct ggg aca gac ttc act ctc acc atc agc -3′ and isolated clones containing human $V_κ$ segments. To identify functional genes we determined the nucleotide sequence of several of the clones and looked for TATA box sequences, open reading frames encoding leader and variable peptides (including 2 cysteine residues), splice sequences, and recombination heptamer-12 bp spacer-nonamer sequences. The light chain construct that we describe in this paper contains a single $V_κ$-III segment ($V_κ$ 65.8) isolated from phage clone 65.8. The sequence of this gene is identical to that of a previously reported germline human $V_κ$ gene that appears to encode the light chain variable sequence of several reported IgM anti-IgG autoantibodies (13). This gene (HUMIGVA27) has been mapped to the Ab region of the human light chain locus (14).

pKC1. We screened a human genomic DNA phage library with κ light chain specific oligonucleotide probes and isolated clones spanning the J-$C_κ$ region. We cloned a 5.7 kb ClaI/XhoI fragment containing $J_κ1$ together with a 13 kb XhoI fragment containing $J_κ2-5$ and $C_κ$ into pGP1d to create the plasmid pKcor. This plasmid contains $J_κ1-5$, the κ intronic enhancer and $C_κ$ together with 4.5 kb of 5′ and 9 kb of 3′ flanking sequences. It also has a unique 5′ XhoI site for cloning $V_κ$ segments and a unique 3′ SalI site for inserting additional cis-acting regulatory sequences. The κ light chain minilocus transgene pKC1 (Figure 1) was generated by inserting a 7.5 kb XhoI/SalI fragment containing $V_κ$ 65.8 into the 5′ XhoI site of pKcor. The transgene insert was isolated by digestion with NotI prior to injection.

Generation of transgenic mice

We isolated the NotI inserts of plasmids pIGM1, pHC1, and pKC1 away from vector sequences by agarose gel electrophoresis. We then microinjected the purified inserts into the pronuclei of fertilized (C57BL/6×CBA)F2 mouse embryos and transferred the surviving embryos into pseudopregnant females as described by Hogan *et al.* (15). We analyzed the mice that developed from injected embryos for the presence of transgene sequences by Southern blot hybridization of tail DNA. We obtained 2 independent lines of mice containing the pIGM1 insert, 12 lines of mice containing the pHC1 insert, and 5 lines containing the pKC1 insert.

Serum analysis

We isolated serum from the blood of transgenic and non-transgenic animals and assayed for the presence of transgene encoded human Igκ, IgM and IgG_1 by ELISA as described by Harlow and Lane (16). Microtiter plate wells were coated with mouse monoclonal antibodies specific for human Igκ (clone 6E1, #0173, AMAC, Inc. Westbrook, ME), IgM (clone AF6,

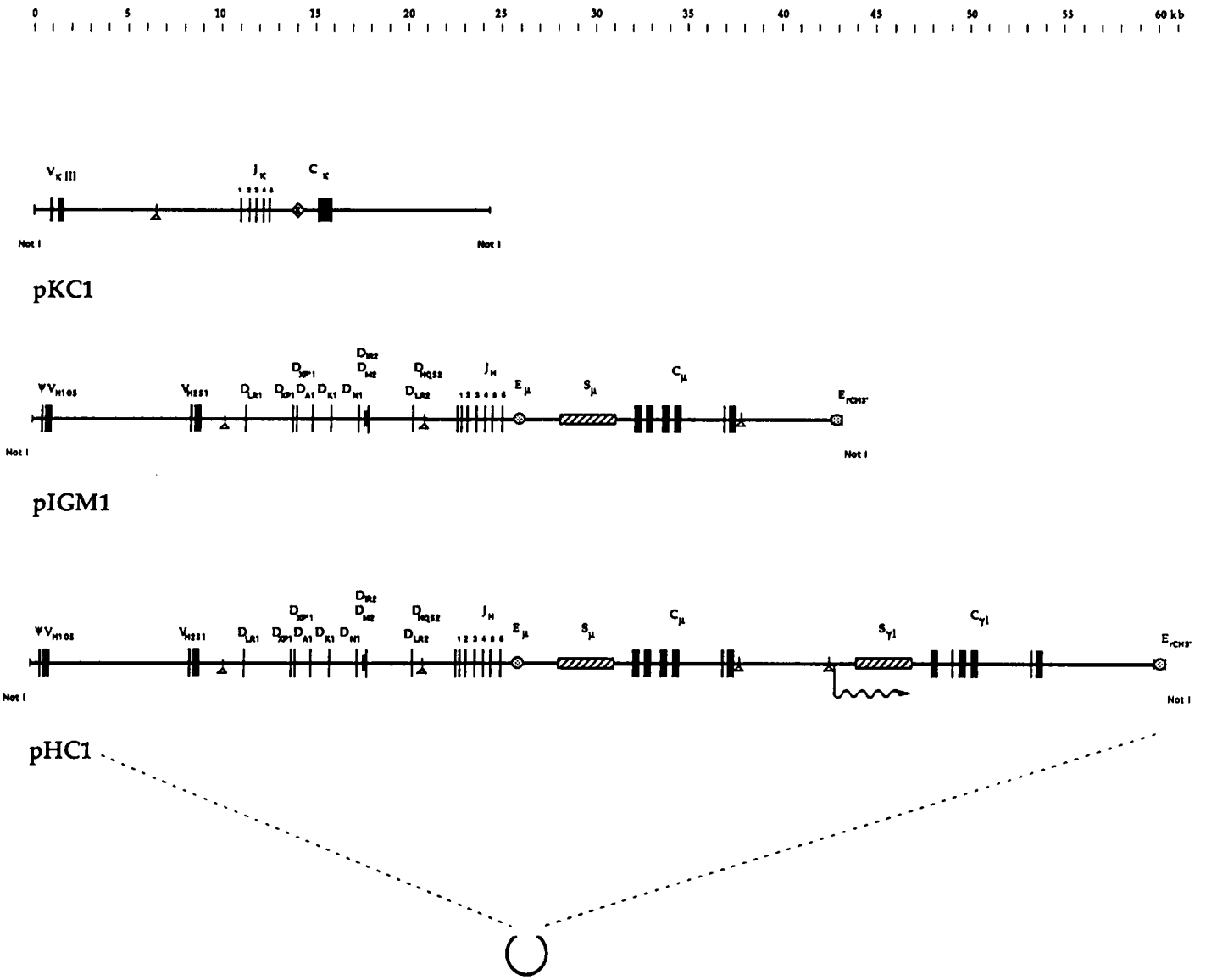

Figure 1. Human immunoglobulin minilocus transgene constructs. The three transgene inserts—KC1, IGM1, and HC1—are depicted as they appear prior to microinjection (after linearization with the restriction enzyme NotI and isolation from vector sequences). The open triangles indicate discontinuities between the structure of the transgene and the natural chromosomal structure of the intact human gene loci. The start site of the human $\gamma 1$ pre-switch sterile transcript is indicated by the wavy arrow below HC1. V, variable segment; D, diversity segment; J, joining segment, C, constant region gene; S, switch region; E, enhancer.

#0285, AMAC, Inc. Westbrook, ME) and human IgG_1 (clone JL512, #0280, AMAC, Inc. Westbrook, ME). Serum samples were serially diluted into the wells and the presence of specific immunoglobulins detected with affinity isolated alkaline phosphatase conjugated goat anti-human Ig (polyvalent) that had been pre-adsorbed to minimize cross-reactivity with mouse immunoglobulins. We used monoclonal human IgG_1,x (#0575, AMAC, Inc. Westbrook, ME) and human IgM (#009-000-012, Jackson ImmunoResearch Laboratories, Inc, West Grove, PA) as standards.

cDNA clones

To assess the functionality of the pHC1 transgene in VDJ joining and class switching we examined the structure of immunoglobulin cDNA clones derived from transgenic mouse spleen mRNA. We isolated pA^+ RNA from the spleens of transgenic mice (**17**) and used this RNA to synthesize oligo-dT primed single stranded cDNA (**18**). The resulting cDNA was then used as template for PCR amplifications using the following synthetic oligonucleotides as primers: VH251 specific oligo-149, cta gct cga gtc caa gga gtc tgt gcc gag gtg cag ctg (g,a,t,c); human $\gamma 1$ specific oligo-151, ggc gct cga gtt cca cga cac cgt cac cgg ttc; and human μ specific oligo-152, cct gct cga ggc agc caa cgg cca cgc tgc tcg. We isolated the resulting 0.5 kb PCR products from an agarose gel, digested with XhoI and cloned the fragments into the plasmid pNNO3. We determined the nucleotide sequences of the inserts by the dideoxy chain-termination method and compiled the data using the GeneWorks sequence analysis software (IntelliGenetics, Mountain View, CA).

Flow cytometry

We prepared single cell suspensions of splenocytes by crushing the spleens between frosted glass slides and lysing the red cells in NH_4Cl (**19**). Lymphocytes were stained with the following

Table 1. Transgenic founder animals generated with the KC1, IGM1, and HC1 miniloci

transgene	line #	~ copy #	serum expression
KC1	665	10–50	κ
	670	1–2	–
	673	1–2	κ
	674	10–50	κ
	676	5–20	κ
IGM1	6	10–50	μ
	15	5–20	μ
HC1	19	1–2	–
	21*	<1–	
	26	5–20	μ, γ_1
	29**	>100	n.d.
	38	5–20	μ, γ_1
	57	10–50	μ, γ_1
	58	10–50	n.d.
	112	1–2	μ
	117	5–20	μ, γ_1
	118	5–20	μ, γ_1
	119	10–50	μ, γ_1
	122	10–50	μ, γ_1

Each transgenic line is designated by the I.D. # of the founder animal that developed from a microinjected embryo. The approximate number of copies of the inserted transgene is estimated by the intensity of the southern blot hybridization signal. Expression of human κ, μ, and γ_1 epitopes was determined for most of the lines by ELISA of serum from either the founder animal or one of its descendants (n.d.: experiment not done; * mosaic; no positive offspring; ** hydrocephalic, died at 6 weeks).

reagents: biotin conjugated anti-human IgM (clone G20-27; Pharmingen, San Diego, CA), FITC conjugated anti-human IgM (clone G20-27; Pharmingen, San Diego, CA), FITC conjugated anti-mouse IgM (clone R6-60.2; Pharmingen, San Diego, CA), biotin conjugated anti-human Igκ (clone G20-361; Pharmingen, San Diego, CA) and Cy-Chrome conjugated anti-mouse B220 (clone RA3-6B2; Pharmingen, San Diego, CA). Biotin conjugated reagents were then stained with phycoerythrin conjugated streptavidin (Becton Dickinson, San Jose, CA). Stained cells were analyzed using a FACscan flow cytometer and LYSIS II software (Becton Dickinson, San Jose, CA). Macrophages and residual red cells were excluded by forward and side scatter.

RESULTS

Human immunoglobulin minilocus constructs

To generate minilocus transgenes we have constructed large plasmid inserts assembled from multiple disparate chromosomal segments. That assembly required serial cloning steps, with the difficulty increasing at each step as the size and sequence complexity increased. To simplify this assembly process, we began by generating a new set of vectors specifically designed for building large transgenes. These vectors (pGP1a, pGP1b, pGP1c, pGP1d, pGP1f, and pGPe) are pBR322-based plasmids that are maintained at a lower copy number per cell than the pUC vectors (**20**). The vectors also include a strong transcription termination signal derived from the trpA gene (**7**). This termination signal should reduce the potential toxicity of coding sequences inserted into the NotI site by eliminating read-through transcription from the ampicillin resistance gene. In addition, these vectors contain polylinkers that are flanked by restriction sites for the rare-cutting enzyme NotI; thus allowing for the isolation of the insert away from vector sequences prior to embryo microinjection.

Figure 1 depicts one light chain minilocus construct and two heavy chain minilocus constructs that we have used to generate transgenic animals. The light chain transgene, pKC1, consists of a single Vκ-III family variable segment, all five human Jκ segments, and the human Cκ segment, together with 8 kb of downstream sequences. The entire 25 kb transgene insert can be isolated using the restriction enzyme NotI. The first heavy chain transgene, pIGM1, consists of a single functional V_H-V family variable segment, ten different human D segments, all six human heavy chain J segments, the μ switch region, and the entire human μ coding region. In addition the construct includes the rat heavy chain 3′ enhancer. The 43 kb transgene insert can be isolated using the restriction enzyme NotI. The final construct, pHC1, is identical to pIGM1 except for the insertion, after the μ gene, of an additional 18 kb of sequence that includes the human γ1 gene and switch sequences. This construct includes the transcription start site for the sterile transcript associated with isotype switching to γ1 (**21**). Like the two other transgene inserts this 61 kb transgene insert can also be isolated using the restriction enzyme NotI. We isolated and microinjected each of these three transgene inserts into mouse embryo pronuclei and generated a total of 19 transgene-positive founder animals (table 1).

Detection of human-sequence immunoglobulins in the serum

We collected serum samples from transgenic and control non-transgenic littermates and looked for the expression of human Igκ, μ, and γ_1 epitopes by ELISA. All of the control non-transgenic mice tested negative for serum expression of human Igκ, μ, and γ_1 epitopes by this assay. Mice from the two lines containing the pIGM1 NotI insert (lines #6 and 15) express human μ. We tested mice from ten lines that contain the pHC1 insert and found that one of the lines (line # 112) expresses low levels of human μ but no detectable human γ_1, seven of the lines (lines #26, 38, 57, 117, 118, 119, and 122) express both human IgM and human IgG1, while mice from two of the lines (lines #19 and 21) do not express detectable levels of human immunoglobulins. Expression levels varied between lines and between individual mice, with mice derived from the multi-copy lines #26 and 57 expressing the highest levels. These mice express human IgM and IgG1 at levels ranging from 0.1 to 1 microgram/ml. Of the three HC1 lines that did not express both transgene encoded isotypes and the one KC1 line that did not express human κ, all were either low copy, mosaic or both. One of these non-expressing lines (#21) was a mosaic that did not pass the transgene on to its offspring. It is possible that transgene-containing cells did not populate the hematopoietic lineage in significant numbers. Two of the lines (#19 and #670) appear by southern blot hybridization intensity to contain only one or two copies of the transgene. The transgene inserts may not be full length in these lines, or the level of expression may be below that which is detectable by our assay. Similarly, line HC1–112, which expresses μ but not γ_1, may be missing 3′ transgene sequences necessary for γ_1 expression.

Cell surface expression of transgene encoded immunoglobulins

We isolated spleen and peripheral blood lymphocytes from eight different lines of transgenic mice; two lines containing the light chain transgene, KC1, and six lines containing the heavy chain

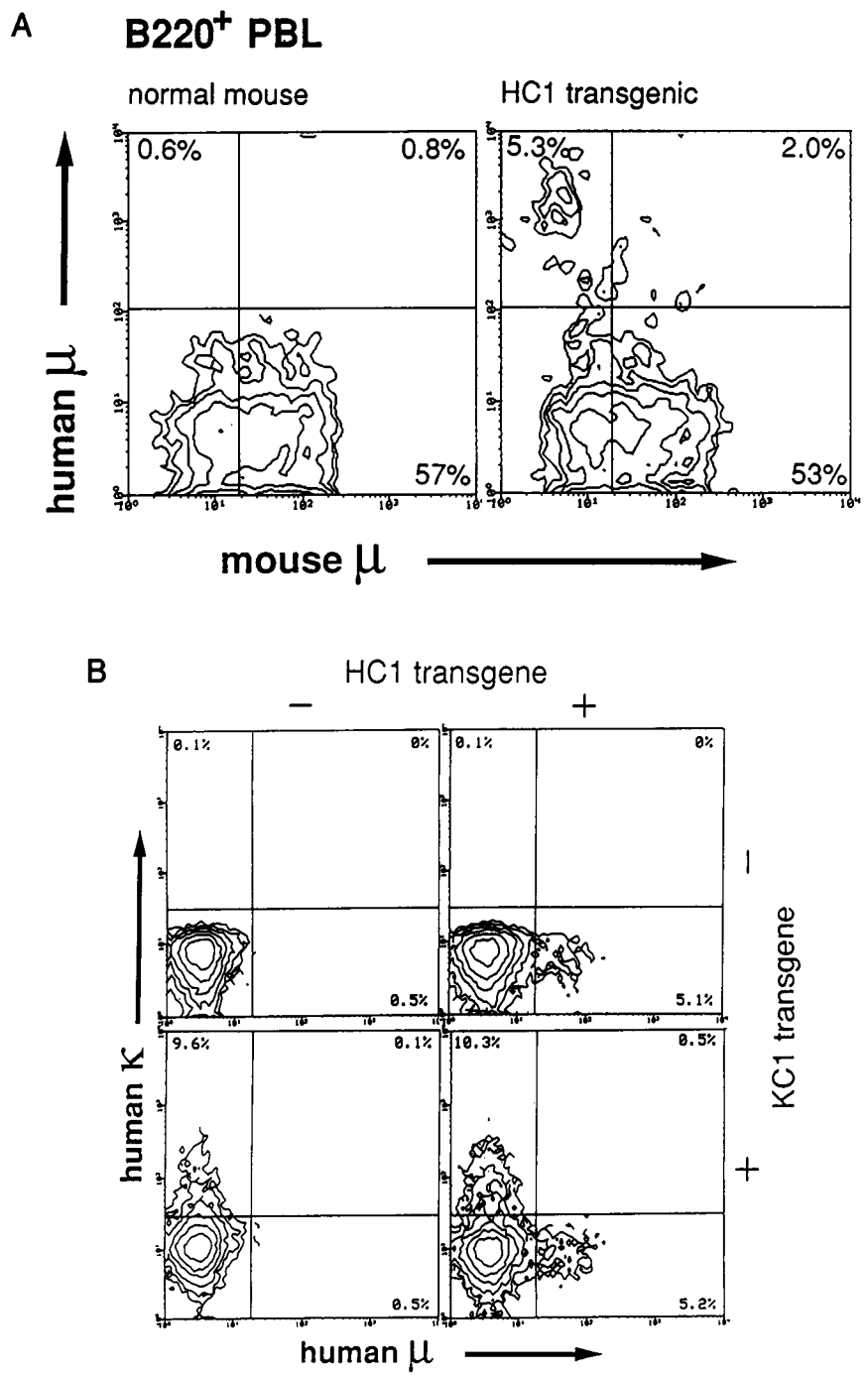

Figure 2. Detection of human sequence immunoglobulins on the surface of transgenic B cells by flow cytometric analysis. (A) Peripheral blood lymphocytes from a negative control and an HC1 line 26 transgenic animal gated for expression of the mouse B cell antigen B220 and assayed for mouse μ (FITC, x-axis) and human μ (PE, y-axis). (B) Spleen lymphocytes from four littermates: upper left, # 1072, HC1-57 negative; KC1-674, negative; upper right, # 1074, HC1-57 positive; KC1-674, negative; lower left, # 1069, HC1-57 negative, KC1-674, positive; lower right, # 1073, HC1-57 positive, KC1-674, positive. Cells were gated for expression of the mouse B cell antigen B220 and assayed for expression of human μ (FITC, x-axis) vs. human κ (PE, y-axis). Cell numbers are indicated by contour lines generated using LYSIS II software (Becton Dickinson, San Jose, CA). Number of cells in each quadrant is given as a percent of the B220 positive/lymphocyte scatter gate.

transgene, HC1. A fraction of the lymphocytes from each of these lines expressed human sequence immunoglobulins on their surfaces as assayed by fluorescent antibody staining and flow cytometry. The percentage of B cells expressing the transgene encoded products varied from 1−2% (for the single copy heavy chain line HC1-112) to 10−20% (for the multi-copy heavy chain line HC1-122). This is illustrated by the example shown in Figure 2A. We isolated peripheral blood lymphocytes from an HC1−26 transgenic animal and a negative littermate, and looked for expression of mouse and human μ heavy chain. In both animals the majority of the peripheral blood B cells express the mouse μ heavy chain; however, a fraction of the cells in the transgenic animal express the human μ chain, and a majority of these cells (5% of the total B220 positive cells) are mouse μ dull

	V_K65.8		J_K	CDR3	FR4
Jκ1	CAG CAG TAT GGT AGC TCA CCT CC		G TGG ACG	TTC GGC CAA GGG	QQYGSSPPWT	FGQG
883-1	--- --- --- --- --- --- --		- --- ---	--- --- --- ---	QQYGSSPWT	FGQG
883-3	--- --- --- --- --- --- --G	A-	- --- ---	--- --- --- ---	QQYGSSPTWT	FGQG
883-4	--- --- --- --- --- --- ---		--- ---	--- --- --- ---	QQYGSSPWT	FGQG
883-11	--- --- --- --- --- --- ---	-	-- ---	--- --- --- ---	QQYGSSPRT	FGQG
883-12	--- --- --- --- ---		---	--- --- --- ---	QQYGST	FGQG
883-13	--- --- --- --- --- --- ---	-	-A ---	--- --- --- ---	QQYGSSPRT	FGQG
883-15	--- --- --- --- --- ---		--- ---	--- --- --- ---	QQYGSSWT	FGQG
883-17	--- --- --- --- --- --- ---	-	-- ---	--- --- --- ---	QQYGSSPRT	FGQG
878-19	--- --- --- --- --- --- --		- ---	--- --- --- ---	QQYGSSPT	FGQG
878-20	--- --- --- --- --- ---		--- ---	--- --- --- ---	QQYGSSWT	FGQG
878-25	--- --- --- --- --- --- --		- ---	--- --- --- ---	QQYGSSPT	FGQG
878-28	--- --- --- --- --- --- ---	-	-- ---	--- --- --- ---	QQYGSSPRT	FGQG
878-30	--- --- --- --- --- --- --		- --- ---	--- --- --- ---	QQYGSSPWT	FGQG
Jκ2	CAG CAG TAT GGT AGC TCA CCT CC		G TAC ACT	TTT GGC CAG GGG	QQYGSSPPYT	FGQG
883-9*	--- --- --- --- --- --- ---	-	--- ---	--- --- --- ---		
883-16	--- --- --- --- --- ---	--C AT	- --- ---	--- --- --- ---	QQYGSSPMYT	FGQG
878-26	--- --- --- --- --- ---	T-- -A	- AG- ---	--- --- --- ---	QQYGSSSQST	FGQG
Jκ3	CAG CAG TAT GGT AGC TCA CCT CC		A TTC ACT	TTC GGC CCT GGG	QQYGSSPPFT	FGPG
883-7	--- --- --- --- --- --- --		- --- ---	--- --- --- ---	QQYGSSPFT	FGPG
883-10	--- --- --- --- --- --- --		- --- ---	--- --- --- ---	QQYGSSPFT	FGPG
883-53	--- --- --- --- --- --- --		- --- ---	--- --- --- ---	QQYGSSPFT	FGPG
878-21	--- --- --- --- --- --- ---			--- --- --- ---	QQYGSSP	FGPG
878-29	--- --- --- --- --- --- ---			--- --- --- ---	QQYGSSP	FGPG
878-31	--- --- --- --- --- --- ---			--- --- --- ---	QQYGSSP	FGPG
Jκ4	CAG CAG TAT GGT AGC TCA CCT CC		G CTC ACT	TTC GGC GGA GGG	QQYGSSPPLT	FGGG
883-2	--- --- --- --- --- --- --		- --- ---	--- --- --- ---	QQYGSSPLT	FGGG
878-27	--- --- --- --- --- --- ---		--- ---	--- --- --- ---	QQYGSSPLT	FGGG
878-32	--- --- --- --- --- --- ---	GG	- GC- ---	--- --- --- ---	QQYGSSPGAT	FGGG
878-34	--- --- --- --- --- --- ---	--	- AG- ---	--- --- --- ---	QQYGSSPPST	FGGG
Jk5	CAG CAG TAT GGT AGC TCA CCT CC		G ATC ACC	TTC GGC CAA GGG	QQYGSSPPIT	FGQG
883-14*	--- --- --- --- --- ---	T	- --- ---	--- --- --- ---		
883-18	--- --- --- --- --- --- ---		--- ---	--- --- --- ---	QQYGSSPIT	FGQG
883-71	--- --- --- --- --- --- ---	--	- ---	--- --- --- ---	QQYGSSPPT	FGQG

Figure 3. Transgene encoded light chain CDR3 sequence diversity. The nucleotide and translated amino acid sequence of the junctional region of 29 independent cDNA clones is shown. Out of frame VJ joints are indicated by asterisks and are not translated. Sequences are divided into categories based on J segment use. The germline encoded sequence is depicted above each category. A dash indicates no divergence from the germline sequence and a blank space or a letter indicates a missing or substituted nucleotide. Each clone is identified by two numbers separated by a dash; the first number indicates the ID # of the animal that provided the RNA, and the second number specifies the clone. Animal #883 was a double (heavy and light chain minilocus) transgenic derived from lines HC1-26 and KC1-665 (heavy chain sequences from this animal are shown in Figure 4). Animal #878 contained only the light chain minilocus (line KC1-665).

or negative. This suggests that the transgene encoded receptor is xenotypically excluding the rearrangement of the endogenous heavy chain gene.

Figure 2B shows FACs profiles of the splenic B cells from four littermates from a cross between a heavy chain transgenic male and a light chain transgenic female. One of the four littermates contained both transgenes and expresses human heavy chain and human κ light chain on splenic B cells. The fraction of cells that simultaneously express both human epitopes (0.5% $\mu+\kappa$) is approximately equal to the product of the fractions that express each epitope individually (5% μ and 10% κ). It therefore appears that the individual transgenes are rearranged and/or selected for independently.

Light chain CDR3 sequences

Figure 3 shows the nucleotide sequences of human κ light chain CDR3 sequences derived from 29 individual cDNA clones from a single transgenic animal. We identified 20 unique cDNA sequences from these 29 clones. Therefore, the expressed human light chains represent a diverse repertoire and not a mono- or oligoclonal expansion of a limited set of rearrangements. All 5 of the human Jκ segments are found to be incorporated into κ chain transcripts, with 45% of the rearrangements using Jκ1. A similar preference for Jκ1 has been reported for endogenous mouse light chain rearrangements (22). The only non-germline encoded sequences in the light chain transcripts occur at the VJ junction in approximately one quarter of the analyzed clones.

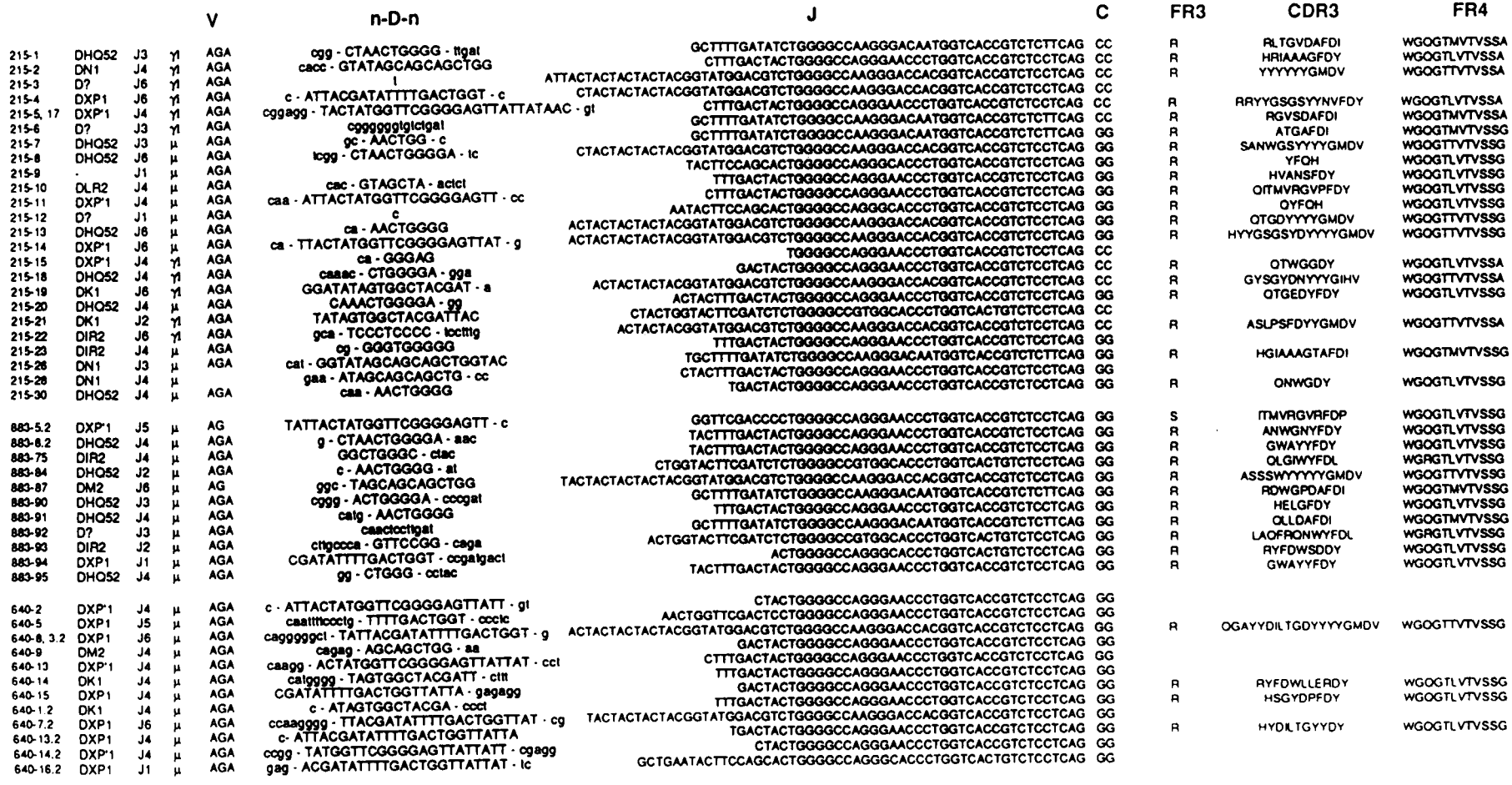

Figure 4. Transgene encoded heavy chain CDR3 sequence diversity. The nucleotide and translated amino acid sequence of the junctional region of 49 independent cDNA clones is shown. Out of frame VDJ joints are not translated. Two of the D segment assignments (clones 215−15 and 883−95) are based on only 5 nucleotides of homology, and therefore represent possible assignments. All other assignments are based on greater than 5 nucleotides of homology. Nucleotides assigned to N regions are in lower case letters. Each clone is identified by two numbers separated by a dash; the first number indicates the ID # of the animal that provided the RNA, and the second number specifies the clone. Animal #215 was a heavy chain minilocus transgenic derived from line HC1-57. Animal #883 was a double (heavy and light chain minilocus) transgenic derived from lines HC1-26 and KC1-665 (light chain sequences from this animal are shown in Figure 4). Animal #640 was a heavy chain minilocus transgenic derived from the single copy line HC1-112. Clones 215-5 and 215-17, and clones 640-8 and 640-3.2 are identical.

Because we do not observe nucleotide changes elsewhere, we interpret these to be the result of random nucleotide additions introduced during VJ joining, and not later somatic mutations occurring during B cell maturation.

Heavy chain CDR3 sequences

Figure 4 shows the nucleotide sequences of human heavy chain CDR3 sequences derived from 49 individual cDNA clones from three different transgenic animals. We identified 47 unique cDNA sequences from these 49 clones. 36 of the 49 clones represented in-frame VDJ joints. This sampling shows that, as observed for the light chain minilocus, the expressed human heavy chains also represent a diverse repertoire and not a mono- or oligoclonal expansion of a limited set of rearrangements. Both μ and $\gamma 1$ sequences are represented. All six human J_H segments are incorporated, and eight of the ten transgene encoded human D segments are found in heavy chain transcripts. Also as observed for the light chain clones, there was no evidence of somatic mutation in the heavy chain sequences. Essentially all of the non-germline encoded nucleotides occurred at V-D, D-J, or V-J junctions and could be ascribed to N region addition. The frequency of non-germline encoded nucleotides outside of N regions is approximately 0.2% (data not shown) and may be the result of errors introduced by reverse transcription and PCR amplification. Because none of the mice had been immunized or exposed to pathogens (all animals were housed in micro isolator cages and were healthy) it is not surprising that we find no evidence of somatic hypermutation.

DISCUSSION

Xenotypic exclusion

We have shown by flow cytometric analysis that most of the human IgM-expressing B cells in our transgenic animals express at most low levels of endogenous mouse IgM. This suggests that correct rearrangement of the human transgene is capable of excluding the rearrangement of the mouse heavy chain locus. Confirmation will require structural analysis of the endogenous loci from a statistically significant number of hybridomas expressing the transgene. However, Nussenzweig *et al.* (**23**) reported exclusion of endogenous μ expression in transgenic mice containing a rearranged human μ gene. Rearrangement exclusion appears to depend on the expression of the transmembrane form of the heavy chain (**24, 25**) and presumably requires that it forms a functional complex with the products of the B29 and mb-1 genes (**26, 27**). Therefore, the xenotypic exclusion implied by our data and that of others suggests that the human heavy chain is capable of forming a functional complex together with the endogenous mouse non-IgH components of the receptor, and that this hybrid complex can induce B cell maturation beyond the developmental stage during which VDJ joining takes place.

Light chain CDR3 sequence analysis

The light chain minilocus encoded transcripts are diverse and incorporate all five human Jκ segments. Approximately one quarter of the Vκ-Jκ joints include non-germline encoded sequences. This addition of junctional random nucleotides is commonly associated with heavy chain N regions (**28, 29**). The large number of naturally occurring Vκ segments makes it difficult to determine whether or not N region addition is a normal component of κ light chain VJ joining (**30 - 33**); however, because the KC1 minilocus construct contains only a single variable segment, the transgenic result is unambiguous. Similar N region additions have been reported previously in light chain transgene rearrangements (**34**). It is possible that the abnormal chromosomal location of the transgene or the concatenated structure of the integrated locus could lead to premature rearrangement accompanied by N region addition. Alternatively, limited N region addition may be a normal component of light chain rearrangement that is difficult to recognize beneath the usual diversity of κ variable segments and somatic mutations. Whether or not the observed light chain N regions are an artifact of the transgenic system, they do not lead to abnormally long CDR3 sequences because the additions are compensated for by exonucleolytic reduction of the V and J segments. Six of the seven transcripts with N region additions result from in-frame VJ joints. Of these, five produce a ten amino acid CDR3 (the expected length given exact V-J joining with no exonucleolytic activity) and the sixth generates a nine residue CDR3. Furthermore, out of all of the 27 in frame transcripts we analyzed, 15% have 8 residue CDR3 sequences while 52% have 9 residue and 19% have 10 residue CDR3's. In comparison, analysis of the 34 naturally occurring Vκ-III nucleotide sequences reported by Kabat (**35**), shows that 12%, 71%, and 15% have 8, 9, and 10 residue CDR3's respectively. Therefore, N region addition does not appear to skew the size distribution of the light chain CDR3's away from that of an authentic human repertoire.

Heavy chain CDR3 sequences

Incorporation of J and D segments. The heavy chain minilocus-encoded transcripts are also diverse and incorporate all 6 human J_H segments, at least 8 of the 10 human D segments, and both heavy chain isotypes included in the transgene. We compared the human heavy chain CDR3 sequences that we isolated from transgenic mice to naturally occurring human CDR3 sequences from published reports (**36, 37**). The transgenic mice preferentially use J_H4 (47%) followed by J_H6 (22%). Yamada *et al.* (**37**) found a similar pattern; 53% of the authentic human joints incorporate J_H4 and 22% incorporate J_H6. It is more difficult to compare D segment usage between the transgenic mice and human PBL because the transgene minilocus does not include all of the human D region. 48% of the 75 in-frame clones analyzed by Yamada *et al.* could be assigned to D segments included in the HC1 transgene, and a further 11% could not be assigned to any known human D segment. These CDR3's either consist almost entirely of N region additions flanking very short D segment remnants or incorporate previously unrecognized D genes. Given these constraints two observations can be made. First, the DXP family is the most heavily used in both the transgenic animals and in human PBL, accounting for 31% and 29% respectively of the in-frame sequences. The second observation is that while only one of the in-frame human PBL sequences used DHQ52, 33% of the in-frame transgenic sequences (25% of all transgenic sequences) used DHQ52.

N region addition. The average length of the CDR3 sequences encoded by the 36 in-frame transcripts from the transgenic animals is 10.6 amino acids. This is similar to the average CDR3 length of 10.3 residues found for adult PBL sequences by Sanz (**36**). However, the transgenic sequences are considerably shorter than the 14.5 residue average found by Yamada *et al.* (**37**) for adult PBL sequences. The length difference between the average naturally occurring heavy chain CDR3 and the sequences found in the transgenic animals is predominantly due to differences in

N region addition. The average number of N region nucleotides per CDR3 sequence (excluding from analysis those sequences for which no D segment could be assigned and thus the N-D border could not be established) is 5.7 for the transgenic sequences and 14.3 for the adult human sequences reported by Yamada *et al.* This average increase in N nucleotides adds approximately 3 amino acids to the authentic human sequences. It appears that 93% of the V-D and D-J junctions in the 88 D containing adult PBL genomic DNA clones reported by Yamada *et al.* include N regions, and that the average length of these individual N regions is 7.7 bp. In contrast, 77% of the heavy chain junctions formed in the transgenic mice include N regions with an average length of 3.8 bp. This is close to the average length of 3 bp/N region for the 63 adult mouse cDNA clones published by Feeney (**38**). Although the heavy chain CDR3 sequences appear superficially like a human fetal liver repertoire because of the overuse of DHQ52 and the shorter average size of the N regions (**39**), the transgenic sequences do not resemble mouse fetal gene rearrangements which are more dramatically reduced in N region addition than human fetal rearrangements. Feeney found that the frequency of N region containing genomic clones fell from 83% in the adult to below 2% in the fetal liver. Therefore, we interpret the fetal character of the CDR3's to be a consequence of mouse B cell N nucleotide addition (which is less extensive than human) coupled with an increase in DHQ52 incorporation that may be peculiar to the transgene.

Implications for the generation of human sequence monoclonal antibodies

If B cells expressing human minilocus-encoded receptors are able to respond to antigen stimulation and undergo affinity maturation it will be possible to use the transgenic animals that we have generated as a source of human sequence monoclonal antibodies. This requires the functional replacement within the B cell receptor complex of the mouse heavy chain by the human heavy chain. It is therefore encouraging that we find B cells in the periphery of transgenic animals that express only the transgene-encoded human heavy chain, indicating that the human/mouse hybrid receptor is able to carry a mouse B cell through development. It is also important that rearrangement of the germline V, D, and J segments generates antibodies that resemble authentic human antibodies. We find that the light and heavy chain CDR3 sequences generated by rearrangement of the introduced miniloci fall within the range of authentic human CDR3 sequences.

ACKNOWLEDGMENTS

We thank Donald Capra and Philip Tucker for discussions and for providing a plasmid clone containing V_H251. We also thank Dennis Huszar, Ted Choi, Manley Huang, Mary Trounstine, Jeanne Loring, Ursula Storb, Frank Costantini, Richard Hardy, and Nila Patil for discussions; and James McCabe for technical assistance. This work was partially funded by NIH grant No. R43 AI31003-01.

REFERENCES

1. Bucchini, D., Renaud, C-A., Ripoche, M-A., Grimal, H., Jami, J. and Weill, J-C. (1987) *Nature* **326**: 409–411.
2. Goodhardt, M. Cavelier, P., Akimenko, M.A., Lutfalla, G., Babinet, C., and Rougeon, F. (1987) *Proc. Natl. Acad. Sci. USA.* 84: 4229–4233.
3. Engler, P., Haasch, D., Pinkert, C.A., Doglio, L., Glymour, M., Brinster, R., and Storb, U. (1991) *Cell* **65**: 1–20.
4. Ferrier, P., Kripple, B., Blackwell, T.K., Furley, A.J.W., Suh, H., Winoto, A., Cook, W.D., Hood, L., Costantini, F., and Alt, F.W. (1990) *EMBO J.* 9: 117–125.
5. Bruggemann, M., Caskey, H.M., Teale, C., Waldmann, H., Williams, G.T., Surani, M.A., and Neuberger, M.S. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**: 6709–6713.
6. Bruggemann, M., Spicer, C., Buluwela, L., Rosewell, I., Barton, S., Surani, M.A., and Rabbits, T.H. (1991) *Eur. J. Immunol.* **21**: 1323–1326.
7. Christie, G.E., Farnham, P.J., and Platt, T. (1981) *Proc. Natl. Acad. Sci. USA.* **78**: 4180–4184.
8. Petterson, S., Cook, G.P., Bruggemann, M., Williams, G.T., and Neuberger, M.S. (1990) *Nature* **344**: 165–168.
9. Yasui, H., Akahori, Y., Hirano, M., Yamada, K., and Kurosawa, Y. (1989) *Eur. J. Immunol.* 19: 1399–1403.
10. White, M.B., Word, C.J., Humphries, C.G., Blattner, F.R. and Tucker, P.W. (1990) *Mol. Cell. Biol.* 10: 3690–3699.
11. Ichihara, Y., Matsuoka, H., and Kurosawa, Y. (1988) *EMBO J.* 7: 4141–4150.
12. Humphries, C.G., Shen, A., Kuziel, W.W., Capra, J.D., Blattner, F.R., and Tucker, P.W. (1988) *Nature* **331**: 446–449.
13. Radoux, V., Chen, P., Sorge, J., and Carson, D. (1986) *J. Exp. Med.* **164**: 2119–2124.
14. Straubinger, B., Huber, E., Lorenz, W., Osterholzer, E., Pargent, W., Pech, M., Pohlenz, H.-D., Zimmer, F.-J., and Zachau, H. (1988) *J. Mol. Biol.* **199**: 23–34.
15. B. Hogan, F. Costantini, and E. Lacy. Methods of Manipulating the Mouse Embryo. (1986) Cold Spring Harbor Laboratory, New York.
16. E. Harlow and D. Lane. Antibodies: A Laboratory Manual. (1988) Cold Spring Harbor Laboratory, New York.
17. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18**: 5294–5299.
18. Gubler, U., and Hoffman, B. J. (1983) *Gene* **25**: 263–269.
19. B. Mishell and S. Shiigi (1980) Selected Methods in Cellular Immunology, W.H. Freeman and Co. New York.
20. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**: 103–119.
21. Sideras, P., Mizuta, T-R., Kanamori, H., Suzuki, N., Okamoto, M., Kuze, K., Ohno, H., Doi, S., Fukohara, S., Hassan, M.S., Hammarström, L., Smith, E., Shimizu, A., and Honjo, T. (1989) *Int. Immunol.* 1: 631–642.
22. Harada K. and Yamagishi, H. (1991) *J. Exp. Med.* 173: 409–415.
23. Nussenzwieg, M.C., Shaw, A.C., Sinn, E., Danner, D.B., Holmes, K.L., Morse, H.C., and Leder, P. (1987) *Science.* 236: 816–819.
24. Manz, J., Denis, K., Witte, O., Brinster, R., and Storb, U. (1988) *J. Exp. Med.* 168: 1363–1381.
25. Nussenzwieg, M.C., Shaw, A.C., Sinn, E., Campos-Torres, J., and Leder, P. (1988) *J. Exp. Med.* 167: 1969–1974.
26. Matsuuchi, L., Gold, M.R., Travis, A., Grosschedl, R., DeFranco, A.L., and Kelly, R. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**: 3404–3408.
27. Reth, M., Hombach, J., Weinands, J., Campbell, K.S., Chien, N., Justement, L.B., and Cambier, J.C. (1991) *Immunol. Today* **12**: 196–201.
28. Lieber, M.R., Hesse, J.E., Mizuuchi, K., and Gellert, M. (1988) *Proc. Natl. Acad. Sci. USA.* 85: 8588–8592.
29. Yancopoulos, G.D., and F.W. Alt. (1986) *Ann. Rev. Immunol.* 4: 339–368.
30. Heller, M., Owens, J.D., Mushinski, J.F., and Rudikoff. (1987) *J. Exp. Med.* 166: 637-.
31. Dersimonian, H., McAdam, K.P.W.J., Mackworth-Young, C., and Stollar, B.D. (1989) *J. Immunol.* 142: 4027–4033.
32. Meindl, A., Klobeck, H.-G., Ohnheiser, R., and Zachau, H.G. (1990) *Eur. J. Immunol.* 20: 1855–1863.
33. Marks, J.D., Tristem, M., Karpas, A., and Winter, G. (1991) *Eur. J. Immunol.* 21: 985–991.
34. Goodhardt, M., Babinet, C., Lutfalla, G., Kallenbach, S., Cavelier, P., and Rougeon, F. (1989) *Nucleic Acids Res.* **17**: 7403–7415.
35. Kabat, E.A., Wu, T.T., Perry, H.M., Gotteman, K.S., and Foeller, C. (1991) Sequences of Proteins of Immunological Interest. Fifth Edition. U.S. Department of Health and Human Services, NIH publication No. 91–3242.
36. Sanz, I. (1991) *J. Immunol.* 147: 1720–1729.
37. Yamada, M., Wasserman, R., Reichard, B.A., Shane, S., Caton, A.J., and Rovera, G. (1991) *J. Exp. Med.* 173: 395–407.
38. Feeney, A.J. (1990) *J. Exp. Med.* **172**: 1377–1390.
39. Schroeder, H.W., Jr., and Wang, J.Y. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87: 6146–6150.