The EMBO Journal vol.9 no.7 pp.2133 - 2140, 1990

Sequence homologies, *N* sequence insertion and J_H gene utilization in V_HDJ_H joining: implications for the joining mechanism and the ontogenetic timing of Ly1 B cell and B-CLL progenitor generation

Hua Gu, Irmgard Förster and Klaus Rajewsky

Institute for Genetics, University of Cologne, Weyertal 121, D-5000 Köln 41, FRG

Communicated by K.Rajewsky

Sequence analysis of rearranged V_HDJ_H genes of B lineage cells from various stages of ontogeny indicates that short sequence homologies at the breakpoints of recombination contribute to *V* region gene assembly. Such homologies are regularly seen at DJ_H junctions of neonatal pre-B cells, most of which do not contain *N* sequences. In the same cells, but not at later developmental stages, preferential usage of the J_H1 element is observed. After birth, *N* sequence insertion increases with time and is always more prominent at the V_HD border than the DJ_H border. In pre-B cells from adult animals and in mature B cells, in cases where *N* sequences were not detectable, sequence homologies at the DJ_H border were found in only half of the instances. This lower incidence could be due to *N* sequence addition to one of the recombining DNA ends and/or cellular selection. Inspection of V_HDJ_H junctions for *N* sequence insertion, sequence homologies at the DJ_H border and J_H1 usage allows the estimation of the timepoint in ontogeny at which particular B cell subsets are seeded into the immune system. Specifically, the present data show that the cells of the Ly1 B cell subset are generated not only neonatally but also beyond the first weeks of life. However, the DJ_H junctions of the progenitors of chronic B cell leukemias which originate from the same B cell subset resemble those of neonatal pre-B cells, suggesting that these cells have already undergone a transforming event at this early developmental stage.

Key words: B cell development/chronic B cell leukemia/gene rearrangements/illegitimate recombination/Ly1 B cells/*N* sequences

Introduction

The B cell compartment of the mouse consists of a variety of B cell subsets which differ from each other in life-style and functional properties. Conventional B cells are continuously generated from stem cells in the bone marrow. Some of these cells are selected by an unknown mechanism into the periphery where they persist for weeks or months as a stable population of resting cells (Förster and Rajewsky, 1990). These cells are thought to be involved in various forms of acquired immunity, in particular, T cell dependent immune responses. Other B cells appear to be seeded into the immune system early in ontogeny and to propagate themselves as mature B cells over long periods of time, perhaps for the lifetime of the organism. One B cell subset of this type is the so-called Ly1 B type of cell which preferentially homes to the peritoneal cavity of the mouse (Herzenberg *et al.*, 1986). Ly1 B cells appear to be involved in functions of natural immunity. Considering the growth properties of these cells, it is not surprising that essentially all chronic B cell leukemias (B-CLL) in mouse and man derive from this B cell subset (for a review of these matters see Kocks and Rajewsky, 1989; Rajewsky *et al.*, 1989).

The growth and selection of B cells *in vivo* is dependent on the specificity of the antibody which they express on their surface. The molecular analysis of antibodies expressed in the various B cell subsets can therefore shed light on the mechanisms by which the cells are selected. This approach has already been successful in elucidating cellular selection in immune responses of conventional B cells and also in the Ly1 B compartment (reviewed by Kocks and Rajewsky, 1989). A limitation of the experiments has been that they were usually based on the immortalization of B cells by the Köhler–Milstein fusion technique (Köhler and Milstein, 1975); this involves selection and requires large numbers of cells. These limitations can be overcome by gene amplification through the polymerase chain reaction (PCR; Saiki *et al.*, 1988). We have used this technique for the construction of cDNA libraries of heavy chain variable (V_H) region genes expressed in pre-B and B cells at various times of ontogeny and in various B cell subsets. The cells were isolated by fluorescence activated cell sorting on the basis of the expression of characteristic cell surface markers.

The genes encoding V_H regions are assembled from V_H, *D* and J_H elements through a process called 'joining' (Alt and Baltimore, 1982; Tonegawa, 1983). From the analysis of the V_H genes rearranged and expressed in the various cellular subsets, we can reach conclusions with respect to cellular selection on the basis of antibody specificity; these results will be presented elsewhere (H.Gu, I.Förster, W.Müller and K.Rajewsky, in preparation). The present paper focuses on the joining process itself. Through its structural footprints in the assembled V region gene, defining the complementarity determining region 3 (CDR3), we show that, like the situation in the compartment of T cells expressing $\gamma\delta$ receptors (Lafaille *et al.*, 1989), the joining process involves the insertion of '*N*' sequences at the recombination breakpoints late, but not early in ontogeny. In the absence of *N* sequence insertion, DJ_H joining is often mediated by short stretches of sequence homology of *D* and J_H at the breakpoint of recombination. In addition, in pre-B cells of the newborn mouse, the J_H1 element is rearranged in preference to other J_H elements. These findings provide the basis for a new way of determining the time in ontogeny at which B cells are seeded into the various B cell compartments and the progenitor cells of B-CLL are generated. In addition, we propose that sequence homology contributes to the joining of V region gene segments in general. This offers a mechanism for the selection of particular recombination breakpoints in the joining reaction on the basis of sequence information in the germline.

Results

V_HDJ_H junctions from pre-B, conventional B and Ly1 B cells at various times of ontogeny

The following subpopulations of B lineage cells were chosen for the analysis of V_HDJ_H joints: pre-B cells from the spleens of 2 day-old animals and from the bone marrow of 4 month-old mice; surface IgM expressing conventional (Ly1 negative) B cells from the spleens of 4 week-old mice, surface IgM and IgD expressing conventional B cells from the spleen and the bone marrow of 4 month-old animals and Ly1 B cells from the spleen of 4 day-old mice and the peritoneal cavity of 4 week-old mice. $1-5 \times 10^5$ cells of each type were isolated by FACS and the purity of the sorted cell populations was determined as described in Materials and methods. Possible contamination with plasma cells was $<2-5 \times 10^{-5}$ (see Materials and methods). RNA was prepared from $\sim 10^4-3 \times 10^5$ cells of each type and cDNA was synthesized using a Cμ-specific primer. The cDNA was poly(dG)-tailed and amplified using a poly(dC)- and a second Cμ-specific primer. Individual V_HDJ_H segments were then isolated by cDNA cloning and sequenced (see Materials and methods for details). Since one aim of our study was to analyze the usage of individual V_H genes in the various B cell subsets, we selected V_HDJ_H segments in which V_H genes of group 1 (Dildrop, 1984) had been rearranged (Gu *et al.*, in preparation).

A limitation of the present approach lies in the fact that V_HDJ_H genes were amplified from cDNA rather than from genomic DNA. This led to the exclusive isolation of productive rearrangements. This selection has to be kept in mind in the interpretation of the data. Our attempts to amplify V_HDJ_H genes from genomic DNA without biasing the analysis with respect to V_H gene distribution have so far been unsuccessful.

The sequences of the various V_HDJ_H junctions are given in Figure 1. In each case, the three most 3′ nucleotides of V_H, the sequence of the D element and the three most 5′ nucleotides of J_H are given. The number of the J_H element used is indicated on the right, together with the designation of the D element and the reading frame in which *Df116* and *Dsp2* elements are translated (Ichihara *et al.*, 1989). In addition, N sequences (i.e. sequences which are not carried in the germline) often appear between V_H and D and between D and J_H. Boxed sequences indicate possible $D-D$ joining events (Kurosawa and Tonegawa, 1982; Meek *et al.*, 1989). The incidence of such sequence patterns was surprisingly high. They include potential $D-D$ joints in which the germline order of the D elements is changed, and joints in which more than two D elements could have participated (Figure 1). $D-D$ joints involving inversion of D elements, as suggested by Shlomchik *et al.* (1990) were not found, however.

In addition to the sequences determined in the present experiments, Figure 1 depicts the junctions of V_HDJ_H segments expressed by Ly1 B cells which had been isolated from 6–10 month-old animals and further propagated *in vivo* as described earlier (Förster *et al.*, 1988), and V_HDJ_H joints from a collection of Ly1 B cell-derived chronic leukemias isolated by Haughton's group (Pennell *et al.*, 1988) and by ourselves (Förster *et al.*, 1988) (Figure 1C).

N sequence acquisition

At the pre-B cell level (Figure 1A), a clear difference emerges between the junctions expressed in cells isolated on day 2 and those from cells of 4 month-old animals: in the latter, N sequences up to 12 bp in length are found almost invariably at the V_H-D border and, in more than half of the cases, between D and J_H. In contrast, in pre-B cells from day 2, N sequences are only rarely inserted and if they are, they are short (maximum length 4 bp). Since pre-B cells are not subject to antigenic selection, the results probably reflect a genuine difference in the joining machinery of 'early' and 'late' pre-B cells, perhaps at the level of terminal deoxynucleotidyl transferase (TdT) expression (Alt and Baltimore, 1982; Rothenberg and Triglia, 1983; Desiderio *et al.*, 1984; Elliott *et al.*, 1988).

As expected, conventional B cells from 4 week- or 4 month-old animals carry V_HDJ_H junctions similar to those carried by pre-B cells from adult bone marrow (Figure 1B). In most cases, N sequences are seen at both the V_H-D and the $D-J_H$ borders, although the latter are shorter on average than the former—a phenomenon observed throughout this analysis.

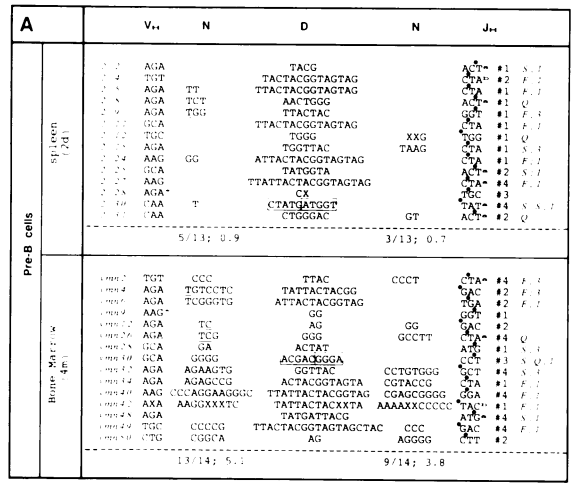
Qualitatively similar data are obtained in the Ly1 B cell compartment; however, N sequence acquisition in the course of ontogeny is slower in this case. V_HDJ_H joints from Ly1 B cells isolated on day 4 carry essentially no N sequences. At 4 weeks, N sequences are observed, but mostly at the V_H-D border; they are still rare between D and J_H. At the age of 6–10 months, the occurrence of N sequences at the V_H-D border is similar to that seen in conventional B cells from week 4 onwards. Roughly half of the cells also carry N sequences between D and J_H. Thus, acquisition of N sequences is also seen in the Ly1 B cell compartment in the course of ontogeny, but it proceeds more slowly than in the compartment of conventional B cells.

Strikingly, very little N sequence is observed at the V_HDJ_H junctions of Ly1-B derived B-CLL: of 11 cases, 6 carry no N sequences at all, 3 have N sequence insertion between V_H and D only and only 2 between both V_H and D and D and J_H. This result suggests that the progenitor cells of B-CLL are often seeded into the immune system early in ontogeny. However, cellular selection cannot be ruled out on the basis of the data in Figure 1C. The results presented in the following sections make this possibility unlikely.

Overlapping sequences of D and J_H mediate DJ_H joining if no N sequence is inserted

A closer inspection of the DJ_H juncitons in Figure 1 reveals a striking phenomenon which is particularly obvious at stages of development where N sequence insertion is rare: DJ_H recombination frequently occurs in positions of sequence overlap. In Figure 2, we show the extent to which this is observed in the V_HDJ_H joints of Figure 1. The DJ_H junctions at which sequence homology betwen D and J_H is involved are depicted, and the relevant D and J_H germline sequences appear above the various joints. Sequence homologies (boxed in Figure 2) range mostly from 2–5 bp. Occasionally, single base pair homologies are seen, and sometimes the stretches of homologies are interrupted by one or two mismatches.

The frequencies at which DJ_H joining occurs in areas of sequence homology in the various cell populations are indicated on the right side of Figure 2. Only junctions lacking N sequences can be considered in this calculation. The results show that in pre-B cells from day 2 (Figure 2A), every single junction follows the homology rule. In pre-B cells of



4 month-old animals, this is true for 3 out of 5 junctions. Since *N* sequence insertion is commonly observed at this stage of development (Figure 1A), it is possible that in cases without obvious sequence homology at the junction, homology was generated by *N* sequences added to one of the recombining DNA ends. Similar frequencies of sequence overlaps at the breakpoint of recombination (in the order of 0.5) are seen in most of the B cell populations, both in conventional (Figure 2B) and Ly1 B cells (Figure 2C). For these cells, *N* sequence addition can only partly explain the lower incidence of joining through homology, as compared to the pre-B cells from day 2. In particular, in the case of Ly1 B cells from 4 day-old animals *N* sequences are not found at the DJ_H border (Figure 1C). Still, only 1 out of 5 joints occurs in areas of sequence overlap. Taking into consideration that B cells, in contrast to pre-B cells, are subject to selection by antigen, we conclude from this result that $D-J_H$ joining occurs predominantly but not exclusively, in regions of sequence overlap, and that cellular selection as a secondary event changes the original pattern of predominance (see also Discussion).

We have also analyzed the V_HD boundaries depicted in Figure 1 for overlaps of the corresponding germline sequences. Only short overlaps of mostly 1 to maximally 3 bases were seen, and these occurred in $<50\%$ of the cases, even at the pre-B cell stage in early ontogeny (not shown). The use of sequence homology in V_HD joining is limited by two factors. One is that stretches of sequence homology are rare at the 3′ end of V_H on the one hand and at the 5′ ends of the various *D* elements on the other. The second reason lies in the conservation of the cysteine in codon 92 of V_H, close to its 3′ end [Kabat *et al.*, 1987; the conservation of this cysteine residue is presumably already required at the pre-B cell stage, because of the interaction of the μ chain with other proteins (Sakaguchi and Melchers, 1986; Kudo *et al.*, 1987; Pillai and Baltimore, 1987)]. This reduces the area in which the joining of V_H to *D* has to occur to a few base pairs, at least in the productive rearrangements analyzed here. In addition, the V_HD joints must be in frame in productive rearrangements—a further limitation in terms of

Fig. 1. V_HDJ_H junctions of rearranged immunoglobulin heavy chain genes expressed in pre-B cells (**A**), conventional B cells (**B**) and Ly1 B cells (**C**). Sequences were either obtained from cDNA clones as described in Materials and methods or taken from the literature ([1]: Förster *et al.*, 1988; [2]: Pennell *et al.*, 1988). Shown are the three most 3′ nucleotides of the V_H gene, the *D* element [identical to known germline *D* elements, Kurosawa and Tonegawa, (1982)], the three most 5′ nucleotides of J_H (with an indication of the reading frame) and *N* sequences at the V_HD and DJ_H borders. In cases where the 3 nucleotides of J_H given in the Figure cannot be assigned unambiguously to the J_H germline sequences, the superscripts indicate the first (a), second (b) or third (c) incidence of these nucleotides (5′ to 3′) in CDR3. 'X' indicates sequence ambiguities. Numbers at the bottom of the various frames indicate the frequency of V_HDJ_H junctions containing *N* sequences (sequences designated by '*' have not been considered in this calculation since the nucleotides between V_H and J_H could not be assigned to either the *D* element or *N* region) and the mean length of *N* regions. Underlined nucleotides in the *N* region may represent *P* sequences (Lafaille *et al.*, 1989). Boxed sequences in the *D* region indicate possible $D-D$ joining. Given on the right hand side of the figure are the *D* elements involved in the junction (Q = *DQ52*; F = *Df116.1*; S = *Dsp2*) and their reading frames (1–3, Ichihara *et al.*, 1989). Circled nucleotides represent differences to the germline sequence. In the case of cDNA clones these nucleotide exchanges may have resulted from PCR amplification.

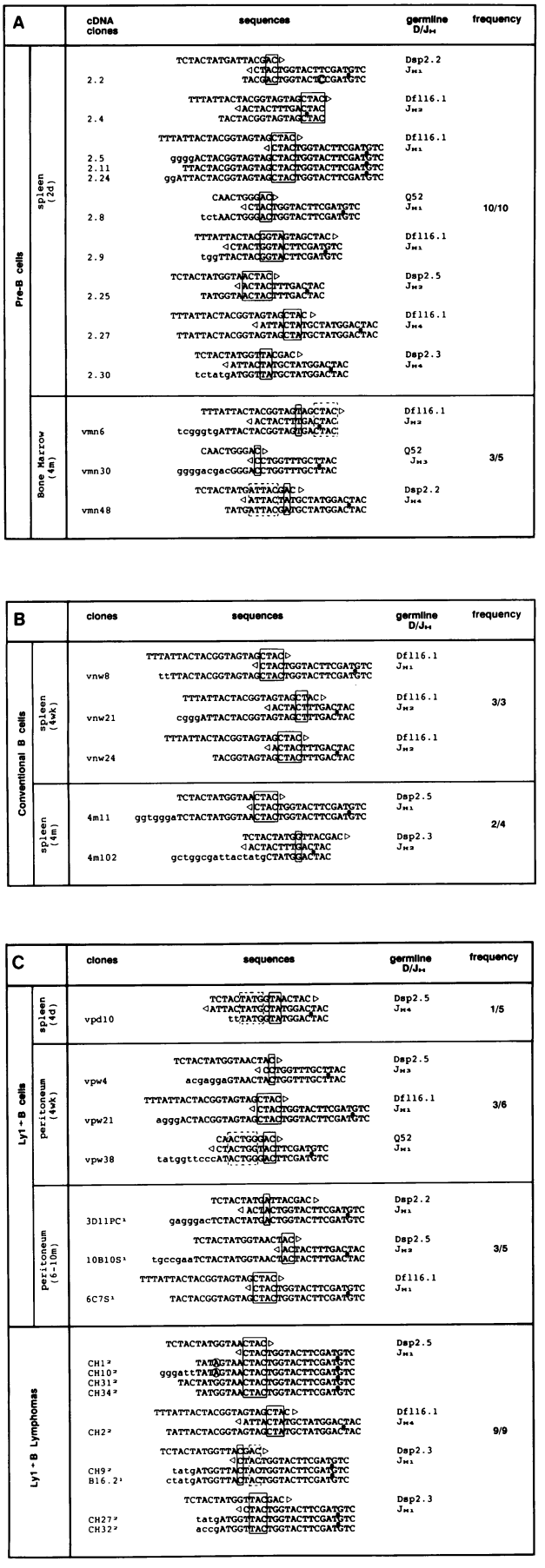


Fig. 2. Short sequence homologies in DJ_H joints. The sequences were derived from different B cell subsets as indicated in the figure (see also legend to Figure 1). Only DJ_H joints with sequence homologies are shown. Above each DJ_H junction, the sequences of the germline D element (ending at the borders of the heptamer signal sequences) and of the 5′ part of the J_H element involved in the rearrangement are depicted. Triangles indicate the position of heptamer signal sequences. Small characters represent N sequences or sequences of another D element in the case of $D-D$ fusions. The reading frame is indicated by a dot in front of the most 3′ codon of CDR3. Circled nucleotides represent mismatches to the known germline D or J_H sequences. The sequences were aligned on the basis of sequence homologies at the recombination breakpoints. Solid and dashed boxes indicate homologies at or close to the region of crossover, respectively. The frequencies shown on the right indicate the number of DJ_H joints containing sequence homologies out of the total number of DJ_H joints with no apparent N sequence insertion.

breakpoints of recombination. It appears therefore, that at the V_H-D border the system avoids the homology rule of joining in order to generate diversity, a situation similar to that of $V\kappa-J\kappa$ joining (see Discussion).

Selection of D element reading frames in pre-B cells and the problem of Dμ proteins

The murine D elements are used in V_HDJ_H genes in all three reading frames, but in the case of *Df116* and *Dsp2* elements, one of the frames, frame I, is strongly preferred in the B cell population (Kaartinen and Mäkelä, 1985; Ichihara *et al.*, 1989), This preference is also seen in the B cell derived V_HDJ_H junctions of our collection (Figure 2B and C). In conventional B cells, reading frame I is used in 26 cases, reading frame II in two cases and reading frame III in one case (Figure 1B). A preference for reading frame I usage is also seen in Ly1 B cells, although at early stages of development, reading frame II is used in one-third of the cases (Figure 1C).

If the overall preference for reading frame I is due to antigenic selection, one would predict that random reading frame usage would be found at the pre-B cell stage. However, as the data in Figure 1A demonstrate, reading frame I is also preferentially used in the (productive) V_HDJ_H junctions of pre-B cells. In order to understand this preference, one has to look at the data in more detail.

Strikingly, not a single pre-B cell junction uses reading frame II. It is this reading frame which allows expression of the so-called Dμ proteins, molecules encoded by the Cμ constant region and *Df116*− or *Dsp*−J_H genes (Reth and Alt, 1984). If cells expressing Dμ proteins have a biological role, one would expect that Dμ protein expression would interfere with further V_H-DJ_H rearrangements. The corresponding V_HDJ_H joints would then be underrepresented in the pre-B cell population, as it is borne out by the data in Figure 1A.

The usage of reading frame III in pre-B cells depends on the D element and N sequence insertion. For *Df116.1*, two stop codons towards its 3′ end limit the usage of reading frame III to cases in which the 3′ portion of the element is lost in the joining process. Significantly, of the three *Df116* elements of this type in the pre-B cell collection (2.9, vmn4 and vmn42), two use reading frame III and only one uses reading frame I. In the case of the *Dsp2* elements, the

Table I. $J_{H}1$ usage in V_HDJ_H rearrangements of B cell subsets at various stages of ontogeny[a]

J_H1	pre-B cells		conv. B cells		Ly1 B cells		
usage	day 2	month 4	week 4	month 4	day 4	week 4	lymphomas
Incidence	8/14	4/15	1/11	3/21	1/6	3/11	8/11
%	57	27	9	14	17	27	73

[a]Calculated from the data in Figure 1.

choice of the reading frame depends on the absence or presence of *N* sequences at the DJ_H border. In their absence, reading frame I is preferred (four out of five cases). We interpret this as a reflection of the promotion of joining by short sequence homologies, which in the case of *Dsp2* favor the usage of reading frame I without exception (cf. Figure 2A). It is noteworthy that this is also true for $Df116-J_H$ joining, except in one case (clone 2.9; the sequence homology in the 2.9 $D-J_H$ junction is probably only rarely used because of its distance from the ends of the recombining DNA strands). In accord with this interpretation, the two *Dsp2* joints in Figure 2A which involve *N* sequences use reading frame III (clones 2.15 and vmn32).

Taken together, the overall preference for reading frame I seen in pre-B cells can be explained by molecular features of the joining process itself (including promotion of joining through sequence homology) and selection against V_HDJ_H joining in cells expressing Dμ protein.

Preferential usage of the J_H1 element in neonatal pre-B cells

A preferential usage of J_H1 is seen in pre-B cells from day 2, but not in any other pre-B and B cell subsets (Table I). It is tempting to think that this may correspond to the preferential usage of *D*-proximal V_H genes in early life (Alt *et al.*, 1987), and that both J_H and V_H gene segments may become available for rearrangement in a developmentally controlled fashion. In principle, the initial preference for J_H1 makes biological sense in allowing diversification in the developing cells through secondary DJ_H rearrangements (Reth *et al.*, 1986; Reth and Leclercq, 1987). However, the validity of this concept remains in doubt as long as so little is understood about life-spans of B lineage cells at developmental stages preceding that of the pre-B cell.

The V_HDJ_H junctions of B-CLLs resemble those of neonatal pre-B cells

The V_HDJ_H junctions of Ly1 B cell derived B-CLL are similar to those of day 2 pre-B cells: *N* sequence insertion is rare, and in its absence, joining occurs in a region of sequence homology in every case (Figure 2C). In addition, the B-CLLs predominantly express the J_H1 element, a phenomenon otherwise exclusively seen in pre-B cells from day 2 (Table I). With respect to DJ_H joining through homology and J_H1 usage, the B-CLLs clearly differ from mature (selected) Ly1 B cells isolated on day 4 after birth (see Figure 2C and Discussion).

Discussion

N sequence insertion and Ly1 B cell generation in ontogeny

The finding that *N* sequences are rarely inserted at V_HDJ_H borders early in ontogeny does not come as a surprise. Similar results have already been obtained in the analysis of T cells expressing γδ receptors (Elliott *et al.*, 1988; Lafaille *et al.*, 1989; Asarnow *et al.*, 1989) and in five hybridomas isolated from newborn BALB/c mice, Holmberg *et al.* (1989) found little *N* region diversity. In the case of DJ_H joining, the load of *N* sequences in cells from adult animals is in general heavier at the V_HD border than at the DJ_H border, and this correlates with an earlier appearance of *N* sequences between V_H and *D* in ontogeny (Figure 1). One can only speculate about the biological function of the ontogenetic timing of *N* region insertion. Clearly, absence of *N* regions implies predominant expression of germline encoded specificities, and this reminds one of the possibility that idiotypic interactions in a germline encoded network may play a functional role in the development of the antibody repertoire early in ontogeny (Vakil and Kearney, 1986; Vakil *et al.*, 1986; Coutinho, 1989; Holmberg *et al.*, 1989). Ly1 B cells represent a major B cell subset at this stage (Hayakawa *et al.*, 1985; Hayakawa and Hardy, 1988), and many of the pre-B cells from day 2 (Figure 1A) may give rise to Ly1 B cells. However, it also becomes clear from the present data that Ly1 B cell generation is not limited in ontogeny to an early phase in which *N* sequence insertion is rare (Figure 1C). This is in accord with a finding by Lalor *et al.* (1989) that Ly1 B cells in normal mice may be generated until 5–6 weeks of age.

A puzzling observation is the usage of reading frame II for the *Df116* and *Dsp2* elements in 4 out of 12 V_HDJ_H joints of Ly1 B cells from 4 day-old and 4 week-old mice. As pointed out in the Results section, reading frame II allows Dμ protein expression upon DJ_H joining, and a frequent usage of this reading frame in early Ly1 B cells could therefore be indicative of some role of Dμ proteins in Ly1 B cell generation. Alternatively, it may represent the footprint of a selection by some unknown antigen.

Sequence homology contributes to variable region gene segment assembly

The most striking observation in the present analysis is the regularity with which DJ_H recombination in pre-B cells occurs in regions of sequence homology. Three such cases were observed in myeloma cells by Kurosawa and Tonegawa (1982), and Alt and Baltimore (1982) drew attention to their possible significance in the context of the joining mechanism. In a recent paper, Ichihara *et al.* (1989) mentioned that they frequently observed short sequence homologies at DJ_H joints lacking *N* sequences. On the basis of the present data we suggest that sequence homology plays an important role in promoting particular joining events. The strongest support for this notion comes from the regularity with which DJ_H joining through sequence homology is observed in pre-B cells. Although these cells might express μ chains on the surface in the absence of conventional L chains (Pillai and Baltimore, 1987) and could thus potentially be selected through the interaction of V regions with external ligands, we think that the regular involvement of sequence homology

in DJ_H joining in pre-B cells reflects a property of the joining process itself rather than cellular selection. The fact that we can explain the pattern of V_HDJ_H rearrangements in terms of *D* region reading frames without having to invoke V region-mediated cellular selection supports this interpretation (see Results). Experimental proof for it could be obtained by looking at non-productive rearrangements in pre-B and also mature B cells. However, so far we have not succeeded in applying the DNA amplification procedure to genomic V_HDJ_H genes.

Recent data in the literature suggest that small stretches of sequence homology may be functionally involved in illegitimate recombination in eukaryotes in general (Roth and Wilson, 1986, 1988; Pfeiffer and Vielmetter, 1988; Thode *et al.*, 1990). The pattern of homologies seen in the present sequence collection (Figure 2) strikingly resembles the patterns observed in the other studies. This suggests, as already proposed by Roth and Wilson (1988) that DJ_H joining as well as other joining reactions of variable region gene segments (see below) follow the general rules of illegitimate recombination once the signal sequences at the borders of these segments have been brought together and the strands have been cut at the base of the signal sequences by a lymphocyte specific recombinase (Alt and Baltimore, 1982). Illegitimate recombination of the coding ends could involve exonucleolytic 'nibbling' and/or double strand invasion by single strand protrusions, alignment of the recombining ends by 'alignment' proteins (which could be part of the recombinase) and joining of the ends mediated preferentially by short sequence homologies (Pfeiffer and Vielmetter, 1989).

A search in the literature reveals that sequence homologies at *VJ* and *VDJ* junctions are widely seen in both T and B lymphocytes. A particularly impressive case is the rearrangements in T cells expressing γδ T cell receptors (Lafaille *et al.*, 1989). The uniform rearrangements in γδ T cells generated early in ontogeny all involve the most appropriate sequence homology at the ends of the recombining gene segments (see Figures 1–5 in Lafaille *et al.*, 1989). This weakens the argument that receptor homogeneity is mediated by 'positive' antigenic selection of such T cells (Lafaille *et al.*, 1990). Rather, the homogeneous joining events are programmed by evolutionary selection of sequence homologies. This of course does not exclude a possible subsequent antigenic selection.

Of special interest is the problem of *VJ* joining in antibody light chain genes, where the extensive diversity seen in CDR3 is mainly due to the imprecision of the joining process (Sakano *et al.*, 1979). Joining homogeneity through sequence homology appears to be avoided in this case in several ways. One is an evolvement of V_L and J_L sequences in the germline such that joining through homology generates an out of frame situation. The extent of 'homologous' productive joining of V_L and J_L is further reduced because, as in the case of V_HD joining (see Results), only short sequence stretches are altogether involved in the joining process, and these exhibit only little sequence homology (Gu and Rajewsky, in preparation). Taken together, the immune system makes use of the promotion of illegitimate recombination by short sequence homologies when homogeneity of receptors is desired (e.g. in certain γδ T cell subsets) and finds ways to avoid this in cases where junctional diversity is required ($V\kappa J\kappa$ and V_HD junctions). In V_HDJ_H rearrangements, *N* sequence insertion also serves the latter purpose.

The generation of B-CLL progenitors in pre-B cells early in ontogeny

The V_HDJ_H joints in the productive heavy chain loci of murine B-CLL resemble those isolated from neonatal pre-B cells in three aspects: the paucity of *N* sequences, the regularity of DJ_H joining through sequence homology and the preferential usage of J_H1. By the same criteria, they are clearly distinguished from all other pre-B and B cell populations which we have analyzed. It is particularly noteworthy that this is also true for the B cell subset to which B-CLL belongs, namely Ly1 B cells. Even early in ontogeny (day 4 after birth) the V_HDJ_H joints of mature Ly1 B cells differ from those of early pre-B cells and of B-CLL in terms of DJ_H joining through sequence homology and J_H1 usage (Figure 1C). This is of importance, since the dominant usage of the same restricted *V* gene repertoire in both Ly1 B cells and B-CLL suggests a similar selection of both types of cells through the interaction of surface immunoglobulin with antigen (Förster *et al.*, 1988; Pennell *et al.*, 1988). Therefore, the most likely interpretation of the difference between V_HDJ_H junctions of B-CLL and those of Ly1 B cells and of their similarity to the junctions of neonatal pre-B cells is that the progenitors of the leukemic cells undergo a first transforming event in the perinatal period, at the pre-B cell stage. (Note in this context that the generation of Ly1 B cells appears in any case to be limited to the first months of life). It is tempting to think that this event is a consequence of the major differentiation process characteristic of pre-B cells, namely *V* region gene assembly. Rearrangements of the immunoglobulin and T cell receptor loci are already known to occasionally involve cellular proto-oncogenes, resulting in tumor development (Cory, 1986; Croce and Nowell, 1986; Adachi *et al.*, 1990). It will therefore be interesting to study the *IgH* and *IgL* loci of B-CLL cells in more detail. A recent study indeed suggests that the *bcl-2* proto-oncogene is linked to the non-productive L chain locus in about 10% of human B-CLL (Adachi *et al.*, 1990).

The postulated early transforming event at the pre-B cell stage is itself presumably not sufficient for B-CLL development. It is typical of B-CLL that its progression is slow and only gradually speeds up with time; the distinction between normal Ly1 B cells and B-CLL is often difficult in murine model systems (Tarlinton *et al.*, 1988). In accordance with general concepts of tumor development (Cooper *et al.*, 1985; Yunis *et al.*, 1987) we assume that in the course of their propagation in the organism, the progenitors accumulate a series of further genetic alterations which finally generates the B-CLL phenotype. Another factor which probably contributes to B-CLL development is the selection of cells through surface immunoglobulin (Pennell *et al.*, 1988; Förster *et al.*, 1988; Kipps *et al.*, 1988). As mentioned above, this selection is evident from the restricted *V* gene usage in B-CLL, which is similar to that seen in the population of normal Ly1 B cells from early ontogeny (Förster *et al.*, 1988 and 1989a). The growth properties of Ly1 B cells (which are long-lived and self-propagating cells) make them particularly suited for leukemogenesis through a cascade of genetic and other factors. In accordance with this view, clones of B-CLL progenitor cells can be identified

in mice long before the manifestation of the disease (Förster *et al.*, 1988).

B-CLL is a frequent form of human leukemia. Considering the similarity of its phenotype with the murine disease, one might speculate that the findings reported here on the ontogenetic origin of murine B-CLL also apply to the human. However, the analysis is more difficult in this case, because most human B lymphocytes appear to carry *N* sequences at V_HDJ_H junctions already in the fetus (Schroeder *et al.*, 1987).

Materials and methods

Mice

All experiments were performed with CB.20 mice which were bred in our own animal facility.

Isolation of cells

B lineage cells used for preparation of cDNA libraries were isolated by fluorescence activated cell sorting using a FACS440. Fluorescence staining of the cells was performed as described previously (Förster and Rajewsky, 1987). For isolation of pre-B cells, spleen cells from a pool of five 2 day-old mice and bone marrow cells from a pool of three 4 month-old mice were stained on the surface with anti-IgM coupled to phycoerythrin (PE) (R33-24-12; Grützmann, 1981) and fluorescein isothiocyanate (FITC)-coupled anti-B220 (RA3.3A1/C1.6; Coffman and Weissman, 1981), and $B220^+IgM^-$ cells were sorted. The lymphocyte population was gated according to orthogonal scatter and forward scatter as described by Förster *et al.* (1989b). To isolate conventional B cells from adult mice, spleen cells from a pool of three 4 month-old mice were stained with anti-IgM^b FITC (MB86; Nishikawa *et al.*, 1986) and biotinylated anti-IgD^b (4/4D7; a gift from Dr T.Tokuhisa) followed by streptavidin–PE. Conventional B cells were sorted as IgM^+IgD^{high} cells in two independent experiments and were used separately for preparation of cDNA libraries. In addition, conventional B cells from the spleens of two 4 week-old mice were sorted as $B220^+Ly1^-$ cells. $Ly1^+$ B cells were isolated from spleen cells derived from fifteen 4 day-old mice or from pooled peritoneal cells derived from ten 4 week-old mice. In these cases staining was performed with anti-Ly1 biotin (53-7.3; Ledbetter and Herzenberg, 1979) followed by streptavidin–PE and anti-B220 FITC. Usually, $1-5 \times 10^5$ cells of each B lineage subset were sorted. Some of these cells were used for reanalysis to determine the purity of the sorted cell populations. The fractions of sorted cells within the desired fluorescence window were 93 and 98% for pre-B cells derived from 2 day- and 4 month-old mice, respectively, 94 and 97% for conventional B cells from 4 week- and 4 month-old mice, respectively, 75% for Ly1 B cells derived from 4 day-old animals, and 97% for Ly1 B cells from 4 week-old mice. $2-5 \times 10^4$ cells from each sorted cell population were transferred to slides, fixed in 95% ethanol, 5% acetic acid and stained in the cytoplasm with anti-IgM FITC to identify plasma cells. No IgM^+ plasma cells were detected among the sorted cells, indicating that possible contamination with plasma cells was below $2-5 \times 10^{-5}$.

V_H gene amplification and cDNA library construction

Total cellular RNA was prepared from $1 \times 10^4-3 \times 10^5$ sorted cells using a direct phenol extraction method: the cells were washed once with phosphate-buffered saline (PBS), pelleted in an Eppendorf tube, and then lysed by adding 30 μl phenol. After vortexing, the cell lysate was extracted by addition of 50 μl extraction buffer [100 mM Tris–HCl (pH 9.0), 100 mM NaCl], 30 μl chloroform and 5 μg tRNA as a carrier. After mixing, phase separation was achieved by centrifugation at room temperature for 3 min. The organic phase was re-extracted with 50 μl extraction buffer. The two aqueous phases were pooled and nucleic acids precipitated overnight in Na-acetate(Ac)/ethanol at −20°C. Genomic DNA was then removed by digestion with DNase I (Boehringer Mannheim) in the presence of an RNase inhibitor (Boehringer Mannheim).

V_H gene amplification was performed according to Loh *et al.* (1989), with modifications. Briefly, first strand cDNA was synthesized with M-MLV reverse transcriptase (BRL) using the Cμ1 primer (see also Förster *et al.*, 1988). The cDNA was successively precipitated with 2.5 M NH_4Ac/ethanol and 0.3 M NaAc/ethanol, and then tailed with poly(dG) according to Dent and Wu (1984). 2 mM $MnCl_2$ was used in the tailing buffer instead of $CoCl_2$ to achieve maximal tailing efficiency. After phenol/chloroform extraction and precipitation in NaAc/ethanol, the tailed cDNA was ready for PCR amplification.

PCR amplification was performed using a Techne thermal cycler. The reaction buffer contained 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 3 mM $MgCl_2$, 400 nM of each dNTP, 10 ng RNase A, 0.01% gelatine, 50 pmol of each primer and 5 U *Taq* polymerase (Perkin Elmer). We used a poly(dC) and a Cμ2 primer containing a *Hin*dIII or *Bam*HI site at the 5′ end, respectively. PCR was performed for 30–35 cycles, depending on the number of cells used in the experiment. Each cycle consisted of 1.5 min heat denaturation at 92°C, 1.5 min primer annealing at 57°C and 2 min primer extension at 70°C. For the first two cycles a lower annealing temperature (47°C) was used. The primer sequences used were: Cμ1: GACAGGGGGCTCTCG (pos. 127–133); Cμ2: CAGGATCC-GAGGGGGAAGACATTTGG (pos. 119–125); poly(dC): GCAAGC-TTCCCCCCCCCC. For construction of cDNA libraries, the PCR product was first cut with *Hin*dIII/*Bam* HI followed by size-fractionation through a 2.5% NuSieve agarose gel (Biozyme) and then cloned into a pTZ19R vector (Pharmacia).

Colony hybridization and DNA sequencing

To identify V_HDJ_H containing clones, a mixed J_H probe ($J_{H1,2}$, 1 kb *Bam*HI fragment, $J_{H3,4}$, 1 kb *Bam*HI/*Xba*I fragment) was used for colony screening. Hybridization was performed as described by Winter *et al.* (1985). Clones containing V_H genes of the J558 V_H gene family (V_H group 1) were identified as described by Rajewsky *et al.* (1989). All probes were labeled with [^{32}P]dATP using a random primer labeling kit (Boehringer Mannheim). V_H gene sequences were obtained by direct plasmid sequencing using the Sequenase™ kit (USB Corporation).

Acknowledgements

We are grateful to U.Weiss, D.Tarlinton, T.Simon, A.Radbruch, C.Kocks, P.Pfeiffer and W.Vielmetter for advice and discussion, and Å.Böhm, G.Schmall and E.Siegmund for typing the manuscript. The work was supported by the Deutsche Forschungsgemeinschaft through SFB 243 and the Fazit Foundation.

References

Adachi,M., Tefferi,A., Greipp,P.R., Kipps,T.J. and Tsujimoto,Y. (1990) *J. Exp. Med.*, **171**, 559–564.

Alt,F.W. and Baltimore,D. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4118–4122.

Alt,F.W., Blackwell,K., Yancopoulos,G.D. (1987) *Science*, **238**, 1079–1087.

Asarnow,D.M., Goodman,T., LeFrancois,L. and Allison,J.F. (1989) *Nature*, **341**, 60–62.

Coffman,R.L. and Weissman,I.L. (1981) *Nature*, **289**, 681–683.

Cooper,M.D., Bertoli,L.F., Borzilla,G.V., Burrows,P.D. and Kubagawa,H. (1985) In *UCLA Symposia on Molecular and Cellular Biology, Leukemia: Recent Advances in Biology and Treatment.* Alan R.Liss, New York, pp. 453–465.

Cory,S. (1986) *Adv. Cancer Res.*, **47**, 189–234.

Coutinho,A. (1989) *Immunol. Rev.*, **110**, 63–87.

Croce,C.M. and Nowell,P.C. (1986) *Adv. Immunol.*, **38**, 245–274.

Dent,G. and Wu,R. (1984) *Methods Enzymol.*, **100**, 96–116.

Desiderio,S.V., Yancopoulos,G.D., Paskind,M., Thomas,E., Boss,M.A., Landau,N., Alt,F.W. and Baltimore,D. (1984) *Nature*, **311**, 752–755.

Dildrop,R. (1984) *Immunol. Today*, **5**, 85–86.

Elliott,J.F., Rock,E.P., Patten,P.A., Davis,M.M. and Chien,Y. (1988) *Nature*, **331**, 627–631.

Förster,I. and Rajewsky,K. (1987) *Eur. J. Immunol.*, **17**, 521–528.

Förster,I. and Rajewsky,K. (1990) *Proc. Natl. Acad. Sci. USA*, in press.

Förster,I., Gu,H. and Rajewsky,K. (1988) *EMBO J.*, **7**, 3693–3703.

Förster,I., Gu,H. and Rajewsky,K. (1989a) In Melchers *et al.* (eds), *Progress in Immunology.* Volume VII, Springer Verlag, Berlin, pp. 389–393.

Förster,I., Vieira,P. and Rajewsky,K. (1989b) *Int. Immunol.*, **1**, 321–331.

Grützmann,R. (1981) Vergleichende idiotypische Analyse von Rezeptoren mit Spezifität für Histokompatibilitätsantigene. Ph.D. Thesis, University of Cologne.

Hayakawa,K., Hardy,R.R., Herzenberg,L.A. and Herzenberg,L.A. (1985) *J. Exp. Med.*, **161**, 1554–1568.

Hayakawa,K. and Hardy,R.R. (1988) *Annu. Rev. Immunol.*, **6**, 197–218.

Herzenberg,L.A., Stall,A.M., Lalor,P.A., Sidman,Ch., Moore,W.A., Parks,D.R. and Herzenberg,L.A. (1986) *Immunol. Rev.*, **93**, 81–102.

Holmberg,D., Andersson,A., Carlsson,L. and Forsgren,S. (1989) *Immunol. Rev.*, **110**, 89–103.

Ichihara,Y., Hayashida,H., Myazawa,S. and Kurosawa,Y. (1989) *Eur. J. Immun.*, **19**, 1849–1854.
Kaartinen,M. and Mäkelä,O. (1985) *Immunol. Today*, **6**, 324–327.
Kabat,E., Wu,T.T., Reid-Miller,M., Perry,H.M. and Gottesmann,K.S. (1987) *Sequences of Proteins of Immunological Interest.* US Government Printing Office, Bethesda, MD.
Kipps,T.S., Tomhave,E., Chen,P.P. and Carson,D.A. (1988) *J. Exp. Med.*, **167**, 840–852.
Kocks,C. and Rajewsky,K. (1989) *Annu. Rev. Immunol.*, **7**, 537–559.
Köhler,G. and Milstein,C. (1975) *Nature*, **256**, 495–497.
Kudo,A., Sakaguchi,N. and Melchers,F. (1987) *EMBO J.*, **6**, 103–107.
Kurosawa,Y. and Tonegawa,S. (1982) *J. Exp. Med.*, **155**, 201–218.
Lafaille,J.J., DeCloux,A., Bonneville,M., Takagaki,Y. and Tonegawa,S. (1989) *Cell*, **59**, 859–870.
Lafaille,J.J., Haas,W., Coutinho,A. and Tonegawa,S. (1990) *Immunol. Today*, **11**, 75–78.
Lalor,P.A., Herzenberg,L.A., Adams,S. and Stall,A.M (1989) *Eur. J. Immunol.*, **19**, 501–506.
Ledbetter,J.A. and Herzenberg,L.A. (1979) *Immunol. Rev.*, **47**, 63–90.
Loh,E.Y., Elliott,J.F., Cwirla,S., Lanier,L.L. and Davis,M.M. (1989) *Science*, **243**, 217–220.
Meek,K., Hasemann,C., Pollok,B., Alkan,S.S., Brait,M., Slaoui,M., Urbain,J. and Capra,J.D. (1989) *J. Exp. Med.*, **169**, 519–533.
Nishikawa,S., Sasaki,Y., Kina,T., Amagai,T. and Katsura,Y. (1986) *Immunogenetics*, **23**, 137–139.
Pennell,C.A., Arnold,L.W., Haughton,G. and Clarke,S.H. (1988) *J. Immunol.*, **141**, 2788–2796.
Pfeiffer,P. and Vielmetter,W. (1988) *Nucleic Acids Res.*, **16**, 907–924.
Pillai,S. and Baltimore,D. (1987) *Nature*, **329**, 172–174.
Rajewsky,K., Gu,H., Vieira,P. and Förster,I. (1989) *Cold Spring Harbor Symp. Quant. Biol.*, **54**, in press.
Reth,M.G. and Alt,F.W. (1984) *Nature*, **312**, 418–423.
Reth,M.G., Jackson,S. and Alt,F.W. (1986) *EMBO J.*, **5**, 2131–2138.
Reth,M. and Leclercq,L. (1987) In Calabi,F. and Neuberger,M.S. (eds.), *Molecular Genetics of Immunoglobulin*, Elsevier Science Publishers B.V. (Biomedical Division), pp. 111–134.
Roth,D.B. and Wilson,J.H. (1986) *Mol. Cell. Biol.*, **6**, 4295–4304.
Roth,D.B. and Wilson,J.H. (1988) In Kucherlapati,R. and Smith,G.R. (eds), *Genetic Recombination.* American Society for Microbiology, Washington DC, pp. 621–653.
Rothenberg,E. and Triglia,D. (1983) *J. Immunol.*, **130**, 1627–1633.
Saiki,R.K.., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 487–491.
Sakaguchi,N. and Melchers,F. (1986) *Nature*, **324**, 579–582.
Sakano,H., Hüppi,K., Heinrich,G. and Tonegawa,S. (1979) *Nature*, **280**, 291–294.
Shlomchik,M., Mascelli,M., Shan,H., Radic,M.Z., Pisetsky,D. and Marshak-Rothstein,A. (1990) *J. Exp. Med.*, **171**, 265–298.
Schroeder,H.W., Hillson,J.C. and Perlmutter,R.M. (1987) *Science*, **238**, 791–793.
Tarlinton,D., Stall,A.M. and Herzenberg,L.A. (1988) *EMBO J.*, **7**, 3705–3710.
Thode,S., Schäfer,A., Pfeiffer,P. and Vielmetter,W. (1990) *Cell*, **60**, 921–928.
Tonegawa,S. (1983) *Nature*, **302**, 79–84.
Vakil,M. and Kearney,J.F. (1986) *Eur. J. Immunol.*, **16**, 1151–1158.
Vakil,M., Sauter,H., Paige,C. and Kearney,J.F. (1986) *Eur. J. Immunol.*, **16**, 1159–1165.
Winter,E., Radbruch,A. and Krawinkel,U. (1985) *EMBO J.*, **4**, 2861–2867.
Yunis,J.J., Frizzera,G., Oken,M.M., McKenna,J., Theologides,A. and Arnesen,M. (1987) *New Engl. J. Med.*, **316**, 79–84.

Received on March 2, 1990; revised on April 2, 1990