The structure of V α and J α segments in the mouse

Jordi Yague⁺, Marcia Blackman, Willi Born¹, Phillipa Marrack¹, John Kappler¹ and Ed Palmer*

Departments of Medicine and Pediatrics and ¹Howard Hughes Medical Institute, National Jewish Center for Immunology and Respiratory Medicine, Denver, CO, USA

Received March 8, 1988; Revised and Accepted October 27, 1988

ABSTRACT

Antigen receptors on most T-cells are heterodimeric glycoproteins, comprised of an alpha chain and a beta chain. These chains are encoded by discontinuous variable (V), diversity (D) and joining (J) gene segments that rearrange to produce a contiguous and functional alpha or beta chain gene. To investigate the size and diversity of the germline repertoire of α -chain gene segments, we have characterized and sequenced 20 alpha chain cDNAs. Among these cDNA clones, we have found 4 J α and 4 V α sequences that have not yet been described. The relationship of these "new" gene segments to those already characterized is discussed.

INTRODUCTION

Antigen receptors on T-lymphocytes are heterodimeric glycoproteins comprised of an alpha and a beta chain (1). The genetic loci encoding these chains are located on chromosomes 6 (beta) and 14 (alpha) of the mouse and each locus consists of non-contiguous gene segments that are brought together by DNA rearrangements to generate a functional T-cell receptor chain gene (2-6). The murine alpha locus consists of approximately 50-100 variable (V) region gene segments, approximately 50 joining (J) region gene segments and a single constant (C) region gene segment (7,8). An unresolved question is whether or not diversity (D) segments are present in the alpha locus. The murine beta chain locus, on the other hand, consists of 21 V region gene segments, 2 D region gene segments, 12 J region gene segments and two constant region genes (9,10).

While the overall structure of the alpha and beta chain loci bears a resemblance to that of the immunoglobulin genes, T-cell receptors and immunoglobulins recognize antigens in ways that are fundamentally distinct. Antibodies are capable of binding to free antigens in solution while T-cell receptors are not. T-cells recognize conventional antigens in physical association with a major histocompatibility complex (MHC) encoded gene product, a phenomenon referred to as MHC-restricted antigen recognition; this is a characteristic property of T-cells. An appreciation of the germline complement of T-cell receptor variable (V) region, diversity (D) region and joining (J) region gene segments in the murine genome is required for a structural understanding of T-cell receptors that

are MHC-restricted. To that end, we have sequenced 20 alpha chain cDNAs isolated from several cDNA libraries. Among these cDNA sequences, we have found 4 new J α and 4 new V α sequences.

MATERIALS AND METHODS

cDNA and Genomic Libraries

cDNA libraries from the hybridomas DO-11.10 and 3DT-52.5 were a generous gift from Dr. Chris Coleclough. These libraries were prepared using a novel technique developed in his laboratory (11) and cloned into the plasmid vector, pUC9. The cDNA library from C57BL/Ka thymocytes was generated by Drs. Howard Gershenfeld and Irv Weissman using the phage vector, λ gt10 (12) while the Charon 4A library of Balb/c genomic DNA was a generous gift from Dr. L. Hood. Libraries were screened using standard procedures (13) using a previously isolated C α probe (14,15).

Subcloning and Sequencing

Plasmid and phage DNAs were prepared using standard protocols; cDNA inserts were excised with the appropriate restriction enzymes and purified by agarose gel electrophoresis. The cDNA inserts were then subcloned into the plasmid vector, pEMBL9⁺ (16) and sequenced using the method of Sanger and Coulson (17). Single stranded template DNA from pEMBL subclones was prepared according to the methods in reference 16. An oligonucleotide primer complementary to the 5' ends of the C α regions was used to initiate the sequencing reactions in one direction while the standard M13 primer was used for sequencing the other strand. We were unable to isolate pEMBL subclones of two cDNAs in both orientations. Sequence information was generated from these two clones by using denatured double stranded plasmid DNA in the sequencing reactions as described previously (18).

Southern Analysis

Various V α DNA fragments were isolated and used to probe murine genomic DNA by the method of Southern as previously described (19). Blots were washed in 0.1X SSC (1X SSC is .15M NaCl, .015M Na citrate) at 50°C.

RESULTS AND DISCUSSION

Alpha chain cDNAs were isolated by screening three different cDNA libraries. One library was derived from C57BL/Ka thymocytes, a second was derived from a T-cell hybridoma, DO-11.10, that recognizes chicken ovalbumin + I-A^d (14), and the third library was derived from a T-cell hybridoma, 3DT-52.5, that recognizes the MHC antigen, D^d (20). From the sequences of 20 T-cell receptor α -chain cDNAs, we have made separate analyses of J α and V α gene segments.

J α genes

The J α sequences shown in Figure 1 demonstrate length heterogeneity, varying

from 19 to 23 amino acid residues. These new J α genes contain the highly conserved sequence, L or V X F G X G T, a feature common to most T-cell receptor and Ig J region genes (21). We have identified another conserved sequence, P X I at the 3' end of these J α genes. The proline residue is encoded by at least 3 of the 4 possible proline codons, while the isoleucine residue is consistently encoded by the triplet, ATC. As the variable residue, X is encoded by a triplet that is either GAC, TAC or AAC, the middle and 3' nucleotides of this codon appear to be invariant. One interpretation consistent with these data is that the variant codon is hybrid in nature. The first nucleotide (G,T or A) is encoded by a particular joining region and the last two nucleotides (AC) are encoded by the C α region in itself. This hybrid codon, apparently formed by RNA splicing, is a feature that is not characteristic of the β chain or Ig loci. The appearance of a hybrid codon on alpha chain cDNAs has also been reported by Arden et al. (7), and has been confirmed by the sequencing of J α and C α genomic DNA (8,22).

Diversity at the V α /J α junction

We have isolated three cDNA clones that contain nearly identical J α sequences (see Figure 2). These sequences are all homologous to J α 1 (8) and differ from one another by only two amino acid residues at the 5' end of these J α genes. Moreover, 5' to this region of variability, the J α -45 cDNA clone contains 251 bp of DNA that include nanomer, spacer and heptamer sequences which probably serve as recognition sites for gene rearrangements while the other two clones have upstream V α sequences. Three hypotheses can explain these nearly identical J α sequences. 1) These three J α sequences may represent three highly related germline gene segments that differ only at their 5'

Figure 1. T-cell receptor alpha chain J region sequences. Four new J α sequences are presented in the figure. These sequences were obtained by sequencing a number of alpha chain cDNAs derived from C57BL/Ka thymocytes (J α 34, J α 42) or Balb/c T-cell hybridomas (J α DO, J α 3DT).

ends. This is unlikely since the rest of the $J\alpha$ nucleotide sequence is identical in these three clones. Even under heavy selection pressure, one might expect there to be some silent mutations resulting from genetic drift. 2) The two variable amino acid residues might be encoded by $D\alpha$ gene segments. cDNA clones 20 and 49 contain $V\alpha$ sequences (not shown in Figure 2) as well as putative $D\alpha$ and $J\alpha$ gene segments, while clone 45 may represent a transcript of a rearranged $D\alpha J\alpha$ complex. It should be noted that these putative D region sequences do not contain a stretch of several contiguous G residues which is characteristic of $D\beta$ and some DH gene segments (23,24). 3) The sequence in clone 45 represents the actual germline sequence of this $J\alpha$ gene segment along with its 5' nanomer-heptamer rearrangement signals. It should be noted that the reading frame upstream from and contiguous with the $J\alpha$ segment in clone 45 is not open; there is no initiation codon and there are two termination codons in this reading frame. The variability seen in clones 20 and 49 would most likely be generated by a mechanism similar to that which generates N-region diversity (25). In this regard N region and D region sequences are conspicuously absent from some α -chain cDNA clones (22,26). Thus, N region diversity and $D\alpha$ segments (if they exist) are not a consistent feature of α -chains.

$V\alpha$ genes

In order to estimate the size and diversity of the germline encoded $V\alpha$ repertoire, we sequenced 20 alpha chain cDNAs. We identified several cDNAs containing constant and joining region gene segments but no variable region sequences. Instead, nanomer/heptamer sequences, presumably signals for DNA rearrangements, as well as 200-300 bp of additional DNA are present immediately 5' to the $J\alpha$ sequences. These cDNAs most likely represent transcripts of unrearranged $J\alpha$ and $C\alpha$ gene segments that are subsequently processed into $J\alpha C\alpha$ mRNAs. Alternatively, these cDNAs could be

Figure 2. The alpha locus may contain D region elements. The sequences of three highly related $J\alpha$ regions is presented in the figure. These $J\alpha$ gene segments differ only at their 5' ends and the differences between these $J\alpha$ sequences may be due to the presence of D elements in the alpha locus or the result of nucleotide insertions by a mechanism similar to that which generates N-region diversity in immunoglobulin genes (see text for discussion).

derived from rearranged DaJa complexes that are transcribed and processed into DJCa mRNAs similar to truncated mRNAs transcribed from the β -chain locus (27) (for discussion of putative Da genes, see above paragraph). We also identified several alpha chain cDNAs containing C region, but neither V nor J region sequences. These clones may represent transcripts of unrearranged, germline C region sequences; similar transcripts have been described at the Ig κ locus (28). Many of the cDNA clones do, however, contain Va sequences. Sequences identical to Va1.2, Va5.1 and Va6.1 were found in this group of clones, as well as four Va gene segments that have not been previously described; the data in Figure 3 display the sequences of these new Va genes. Three of the four new sequences (Va4.4, Va7.2 and Va8.4) fall into existing Va gene

Figure 3. T-cell receptor alpha chain V region sequences. Four new Va sequences are shown in the figure. These sequences were obtained by sequencing a number of alpha chain cDNAs derived from C57BL/Ka thymocytes (Va 7.2, Va 8.4) or BALB/c T-cell hybridomas (Va 4.4, Va 13.1) or a BALB/c germline genomic library (Va 13.2).

families as defined by Arden et al. (7). The V_α proteins within a family are generally 70-90% homologous to each other. During the completion of this work, Chou et al (26) have reported a sequence, $V_\alpha F3.5$ that is identical to $V_\alpha 8-4$. We have isolated a fourth V_α gene ($V_\alpha 13.1$) that does not fall into any existing alpha chain gene family. This V_α gene encodes part of the receptor on the T-cell hybridoma, DO-11.10. We have also isolated a highly homologous V_α gene, $V_\alpha 13.2$ from a BALB/c germline genomic library, the sequence of which is also presented in Figure 3. Surprisingly, over a region of 274 nucleotides, the $V_\alpha 13.1$ and $V_\alpha 13.2$ sequences differ at only four nucleotide positions, three of which encode amino acid differences.

The Southern analyses in Figure 4 demonstrate that these V_α genes are members of multi-gene families. The $V_\alpha 4$, $V_\alpha 7$ and $V_\alpha 8$ families contain 5, 8 and 6 homologous sequences, respectively (Figure 4 and data not shown). The $V_\alpha 13$ family is small in comparison with most other V_α gene families in that only two genomic restriction fragments are detected with a $V_\alpha 13$ probe.

Figure 4. Analysis of V_α gene families in murine genomic DNA. Liver DNA from C57BL6/J mice was digested with Pvu2 or Hind3, electrophoresed through 0.7% agarose gels, transferred to nitrocellulose (19) and probed with various V_α probes. Size markers are indicated for each Southern analysis and the location of C_α containing fragments are indicated in those instances where the probe contained C_α as well as V_α sequences.

Sequence variability within the V α 4 gene family

The data in Figure 5 present a sequence comparison of V α genes within the V α 4 family. This kind of analysis allows one to examine related sequences for regions of variability. The nucleotide and amino acid sequences of V α 4.4 (this paper) are compared with those of V α 4.3, V α 4.2 and V α 4.1 (7,29). The ratio of replacement (R) to silent (S) mutations is low (V α 4.4 vs. V α 4.3 - R/S = 0.85; V α 4.4 vs. V α 4.2 - R/S = 1.5; V α 4.4 vs. V α 4.1 - R/S = 1.18), indicating an overall conservation of the amino acid sequence of the V α 4 domains. Nevertheless, there are three regions of variability indicated in the figure located at amino acid positions 15-19, 25-30 and 49-56 that most likely encode complementarity determining regions (CDR) in the alpha chain protein. We find it noteworthy that the analysis of V α genes within a family identifies fewer regions of variability than does a similar analysis of pooled sequence data from all known V α sequences (see Figure 5 in reference 29). A similar conclusion has been reached by Chou et al (26) in their analysis of six V α 8 germline genes. Comparing the nucleotide and predicted amino acid sequences of the V α 8 genes reveals a single region of hypervariability extending from amino acid position 26 - 33, which is close to a region of

Figure 5. Sequence comparison of four members of the V α 4 gene family. The sequence of V α 4.4 (described in Figure 3) is compared with the sequences of three other members of the V α 4 family. V α 4.1 and V α 4.2 have been described by Arden et al. while V α 4.3 has been described by Becker et al (29).

hypervariability (position 25-30) that we have identified in the V α 4 family (see Figure 5). Both our analysis and that of Chou et al (26) demonstrate the point that regions of hypervariability may be overestimated when one analyses pooled sequence data from all known V α sequences.

Another striking feature of the V α 4 gene family is the absence of variability around position 90, which in the immunoglobulin light chains encodes CDR3 (21). In fact, there is a paucity of amino acid variability among these genes from position 57 to position 94, which contrasts with Ig V κ genes, even those within a family (21). An additional CDR in the alpha chain could be separately encoded by J region sequences given the large germline repertoire (50) of diverse J α genes. This would be similar to Ig heavy chains where CDR3 is separately encoded by D gene segments (21).

The repertoire of Ig receptors is extensive primarily due to a large number of germline gene segments (150-1000 V H , 250 V κ) (30,31,32) and due to the high rate of somatic mutation occurring within and around rearranged Ig V region genes. There are relatively fewer T-cell receptor V gene segments (50 V α , 22 V β) (7,33,34,35,36), more T-cell receptor J gene segments (50 J α , 12 J β) (8,9) in the germline but no obvious high rate of somatic mutation (although this latter point has not been examined thoroughly). In addition, T-cells undergo selection in the thymus where 99% of developing T-cells die intrathymically; this may decrease the size of the peripheral T-cell repertoire by two orders of magnitude. A smaller repertoire of T-cell receptors may still be well suited for antigen recognition if one considers the fact that T-cells recognize antigenic fragments physically complexed with MHC molecules (37,38,39). If these antigen/MHC complexes are alterations of a basic MHC structure, then the universe of T-cell antigens (altered-self MHC) may be limited and a relatively small number of T-cell receptors may be sufficient to carry out specific antigen recognition.

ACKNOWLEDGEMENTS

The authors appreciate the assistance of Ms. Barbara Hill and Ms. Judy Franconi in the preparation of the manuscript. This work was supported by NIH grants AI22259 and AM37070. M. Blackman was supported by a post-doctoral fellowship from the Cancer Research Institute. J. Yague was partially supported by a fellowship from CIRIT, Generalitat de Catalunya, Spain.

*To whom correspondence should be addressed

+Present Address: Department of Immunology, Hospital Clinico, Barcelona, Catalunya, Spain

References

1. Marrack, P., C. Hannum, M. Harris, K. Haskins, R. Kubo, M. Pelanne, R. Shimonkevitz, J. White, and J. Kappler. (1983) *Immunol. Rev.* 76: 131.
2. Caccia, N., M. Kronenberg, D. Saxe, R. Haars, G. Bruns, J. Gorman, M.

- Malissen, H. Willard, Y. Yoshikai, M. Simon, L. Hood, and T. Mak. (1984) *Cell* 37: 1091.
3. Dembic, Z., W. Bannworth, B.A. Taylor, and M. Steimetz. (1985) *Nature* (London) 314: 217.
4. Hedrick, S., D. Cohen, E. Nielsen, and M. Davis. (1984) *Nature* 308: 149.
5. Chien, Y., D. Becker, T. Lindsten, M. Okamura, D. Cohen, and M. Davis. (1984) *Nature* 312: 31.
6. Saito, H., D. Kranz, Y. Takagaki, A. Hayday, H. Eisen, and S. Tonegawa. (1984) *Nature* 312: 36.
7. Arden, B., J.L.Klotz, G. Siu, and L.E. Hood. (1985) *Nature* (London) 316: 783.
8. Winoto, A., S. Mjolsness, and L.E. Hood. (1985) *Nature* (London) 316: 832.
9. Gascoigne, N., Y. Chien, D. Becker, J. Kavalier, and M. Davis. (1984) *Nature* 310: 387.
10. Malissen, M., K. Minard, S. Mjolsness, M. Kronenberg, J. Goverman, T. Hunkapiller, M.B. Prystowsky, Y. Yoshikai, F. Fitch, T.W. Mak, and L. Hood. (1984) *Cell* 37: 1101.
11. Coleclough, C., and F.L. Erlitz. (1985) *Gene* 34: 305.
12. Young, R.A., and R.W. Davis. (1983) *Proc. Natl. Acad. Sci. USA* 80: 1194.
13. Maniatis, T., E.F. Fritsch, and J. Sambrook. (1982) *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
14. Yague, J., J. White, C. Coleclough, J. Kappler, E. Palmer, and P. Marrack. (1985) *Cell* 42: 81.
15. Born, W., J. Yague, E. Palmer, J. Kappler, and P. Marrack. (1985) *Proc. Natl. Acad. Sci. USA* 82: 2925.
16. Dente, L., G. Cesareni, and R. Cortese. (1983) *Nucl. Acids. Res.* 11: 1645.
17. Sanger, F., S. Nicklen, and A.R. Coulson. (1977) *Proc. Natl. Acad. Sci. USA* 74: 5463.
18. Chen, E.Y. and P. H. Seeburg. (1985) *DNC* 4: 165.
19. Wahl, G.M., M. Stern, and G.R. Stark. (1979) *Proc. Natl. Acad. Sci. USA* 76: 3683.
20. Endres, R., P. Marrack, and J. Kappler. (1983) *J. Immunol.* 131: 1656.
21. Kabat, E.A., T.T. Wu, H. Bilofsky, M. Reid-Miller, and H. Perry. (1983) *Sequences of proteins of immunological interest*. U. S. Department of Health and Human Services.
22. Hayday, A., Diamond, D., Danigawa, G., Heilig, J., Folsom, V., Saito, H. and Tonegawa, S. (1985) *Nature* 316: 828.
23. Kavalier, J., M.M. Davis, and Y. H. Chien. (1984) *Nature* 310: 421.
24. Siu, G., M. Kronenberg, E. Strauss, R. Haars, T.W. Mak, and L. Hood. (1984) *Nature* 311: 344.
25. Alt, F.W. and D. Baltimore. (1982) *Proc. Natl. Acad. Sci. USA* 79: 4118.
26. Chou, H.S., Behlke, M.A., Godambe, S.A., Russell, J.H. Brooks, C.G. and Loh D.Y. (1986) *EMBO Journal* 9: 2149.
27. Kronenberg, J., et al. (1985) *Nature* (London) 313: 647.
28. Schibler, U., K. B. Marcu, and R.P. Perry. (1978) *Cell* 15: 1495.
29. Becker, D., Patten, P., Chien, Y. H., Yokota, T., Eshhar, Z., Giedin, M., Gascoigne, N.R.J., Goodenow, C., Wolf, R., Arai, K. I., and Davis, M.M. (1986) *Nature* 317: 430.
30. Broder, P., S. Thompson, and R. Riblet. (1984) In, *Regulation of the Immune System* H. Cantor, L. Chess, and E. Sercarz, eds. Alan R. Liss, New York.
31. Cory, S., B.M. Tyler, and J.M. Adams. (1981) *J. Mol. App. Gen.* 1: 103.
32. Livant, D., Blatt, C., Hood, L. (1986) *Cell* 47: 461.
33. Patten, P., T. Yokata, J. Rothbard, Y. Chien, K. Arai and M. Davis. (1984) *Nature* 312: 40.
34. Barth, R., B. Kim, N. Lan, T. Hunkapiller, N. Sobieck, A. Winoto, H. Gershenfeld, C. Okada, D. Hansburg, I. Weissman and L. Hood. (1985) *Nature* 316: 517.
35. Behlke, M.A., H.S. Chou, K. Huppi and D.Y. Loh. (1986) *Proc. Natl. Acad. Sci. USA* 83: 767.

36. Behlke, M.A., D.G. Spinella, H.S. Chou, W. Sha, D.L. Hartl, and D.Y. Loh. (1985) *Science* 229: 566.
37. Babbitt, B., Allen, P., Malsueda, G., Haber, E., and Unanue, E. (1985) *Nature* 317: 359.
38. Buus, S., Sette, A., Colon, S., Jenis, D., and Grey H. (1986) *Cell* 47: 1071.
39. Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennet, W.S. Strominger, J.L., and Wiley, D.C. (1987) *Nature* 329: 512.