

MINI REVIEW

The genomic organization and complete sequencing of the human T-cell receptor β locus

Geraldo A.S. Passos Jr.

In 1996, human genome research was marked by a milestone from Leroy Hood's group who published the largest contiguous human DNA sequenced to date, i.e., the 685 kb of DNA of the T-cell receptor (TCR) β locus (Rowen *et al.*, 1996). This achievement resulted from a combination of physical mapping and automated DNA sequencing methodology and the correct previous choice of DNA region to be analyzed. The human TCR β locus spans about 0.7 Mb, the appropriate size for current large-scale DNA sequencing. Moreover, this locus plays a vital role in immunity.

T-cell receptors are heterodimeric polypeptides expressed on the surface of T-cells held together by covalent disulfide bonds. TCR α/β is present in both T-helper cells (CD4+) and T-cytotoxic (CD8+) and TCR γ/δ is expressed in a minority of lymphocytes (10% of circulating lymphocytes) that recognize *Mycobacterium*, tumor cells and stressed cells (Raulet, 1989; Kabelitz, 1992). Similar to an immunoglobulin molecule, the TCR chains contains a variable region that interacts with the antigen (amino terminal) and a constant region (carboxy terminal) attached to the cell membrane. The recognition of a non-self antigen by TCR α/β is made in association with the self major histocompatibility complex (MHC) molecules expressed by antigen-presenting cells (APC) such as macrophages,

monocytes, B-cells and dendritic cells. T-cells do not recognize a soluble antigen directly but only those present on the surface of APC or target tumor cells (Kersh and Allen, 1996). The essential role of the immune system is the discrimination between the self and non-self antigens. So, TCR α/β represents a "conceptual molecule" of current immunology, since it recognizes simultaneously the self (MHC) and the non-self (antigen). Each individual T-cell expresses a clonally distributed TCR. The constant region of a TCR chain is coded by a C gene and the variable region results from the junction of non-contiguous DNA segments (V gene and J segment for α and γ chains; V gene, D-J segments for β and δ chains). The maturation of a T-cell depends on the somatic recombination V-J and V-D-J DNA segments assembling an active TCR gene. The V(D)J recombination is mediated by the recombinases RAG-1 and RAG-2 (Oettinger *et al.*, 1990; Mombaerts *et al.*, 1992). The generation of the diversity in TCR is based on i) the multiple germline V, D and J elements, ii) combinatorial joining of these elements during T-cell maturation, iii) N diversity, i-e, the addition of non-germline nucleotides to the gene segments by terminal deoxynucleotidyl transferase, and iv) the combinatorial heterodimeric associations between α and β chains.

Organization of the TCR β locus

The human TCR β locus maps to chromosome 7q35 (Isobe *et al.*, 1985) and comprises about 65 V β genes. Until recently, the data contributing to the partial organization of the TCR β family were derived from the cDNA analysis of ~270 different β transcripts (Con-

Departamento de Morfologia, Disciplina de Genética, Faculdade de Odontologia de Ribeirão Preto, Universidade de São Paulo, 14049-904 Ribeirão Preto, SP, Brasil. Fax: (016) 633-0999, E-mail: passos@highnet.com.br c/c to gadsjun@usp.br and Grupo de Imunogenética Molecular, Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, USP, 14040-900 Ribeirão Preto, SP, Brasil.

cannon *et al.*, 1986; Tillinghast *et al.*, 1986; Kimura *et al.*, 1987) and from a few cloned germline DNAs of the β locus (Li *et al.*, 1991; Slightom *et al.*, 1994; Zhao *et al.*, 1994). The V β gene segments have been divided by cDNA analysis into 26 subfamilies, whose members exhibit 75% DNA sequence homology. cDNA analysis has provided only a partial idea of the genomic organization because it analyzes only the active genes. Thus, a limited picture for the TCR β locus emerged as follows; 5'-(unknown number and order of V β elements)-(D β 1-J β 1.1-1.6-C β 1-D β 2-J β 2.1-2.7-C β 2-V β 20). Today, with the complete sequence of the 685-kb TCR β locus (Rowen *et al.*, 1996) we can have insights into the organization, evolution and diversification of this gene cluster. Eighty-one TCR elements and two other non-TCR genes lie in this sequence. One is the dopamine- β -hydroxylase-like gene at the 5' end of this sequence and eight trypsinogen genes divided into two clusters, three immediately 3' to the dopamine- β -hydroxylase-like gene and five immediately 5' to the D β 1 gene segment. Dot matrix sequence comparisons in relation to TCR cDNA sequences identified 65 V β genes, six of which were new. With one exception all of these genes are located between the dopamine- β -hydroxylase-like gene and the D β 1 gene segment. Each of the duplicated DJC clusters, separated by 2.5 kb, contains one D β and six or seven J β gene segments and a C β gene. An enhancer element is located at the 3' extremity of the C β 2 gene and the 65th V β gene has an inverted translation reading frame compared to the other TCR elements. Of the 685 kb of the locus 4.6% are coding regions. A new nomenclature was proposed based on the complete sequence. The V β subfamilies were assigned consecutively, numbered starting at the 5' end of the locus and the individual subfamily members were then numbered sequentially after the subfamily designation. A β locus translocation from chromosome 7 to chromosome 9 was also observed. The segment represents a duplication and translocation of a DNA segment from the 3' end of the β locus that includes at least seven V β elements and a functional trypsinogen gene denoted T9. The physical map of the human TCR β locus is also available via Internet: <http://imgt.cnusc.fr:8104> or <http://www.ebi.ac.uk/imgt>.

Characterization of the V β gene segments

A V β gene contains five regions: i) a promoter, ii) a first exon coding for a signal peptide, iii) an intron with 5' and 3' RNA splicing signals, iv) a second exon encoding the V element and v) a DNA rearrangement signal sequence. A clustering of the members of individual V β subfamilies was revealed by phylo-

genetic analysis using > 75% sequence homology as a criterion. The 65 V β gene segments are divided into 46 functional genes, 19 pseudogenes, and 22 other additional sequences with limited similarity to V β gene segments were identified, each bearing several major lesions other than pseudogenes. Although these sequences provide no functional information, they contribute to a dynamic view of the evolutionary changes at this locus. The V β gene segments contain conventional RNA splicing signals 5' GT and 3' AG for the introns. The TCR β DNA rearrangement signals are like the other gene segment counterparts (Hesse *et al.*, 1989; Koop *et al.*, 1994), such as human and mouse immunoglobulin, with the classical heptamer-spacer-nonamer structure. The expression of the V β genes shows a heterogeneous pattern, with some V β and J β gene segments being expressed more frequently than others. These differences could arise from clonal selection, different promoter strengths, half-life of the mRNAs, or specific DNA rearrangement probabilities. The expression insights were obtained by comparison of TCR germline and cDNA sequences. Moreover, the cDNA analysis was useful for establishing exon-intron boundaries and for identifying non-TCR genes within the TCR locus, such as the dopamine- β -hydroxylase-like and trypsinogen genes.

Evolution of the locus itself

The β locus presents a high degree of complexity, 47% of its sequences being composed of locus-specific repeats (homology units) that have been duplicated 2 to 10 times, with eight major locus-specific repeats across the multigene family, as deduced from dot-matrix analysis.

Variations in the sequence

The polymorphism data at the TCR β locus (as RFLPs) have been obtained from limited cDNA and germline DNA analyses. Overlapping cosmid clones originated from different haplotypes (chromosomes) showed discrepant rates of sequence variation. Two large insertion-deletion polymorphisms affecting three V β [V β 13S2 (6-2), V β 7S2 (4-3) and V β 9S2 (3-2)] and two trypsinogen genes (T6 and T7), respectively, exhibited allele frequencies of 0.37 (insertion) and 0.61 (deletion) for V β polymorphism and 0.54 (insertion) and 0.46 (deletion) for the trypsinogen polymorphisms. The V β deletion can be associated with loss of autoimmune tendencies, but it is not clear why the loss of the functional trypsinogen T6 gene would confer a selective advantage.

Concluding remarks and perspectives

The last year was marked by a historical achievement in the human genome program: the TCR β locus was totally sequenced, its 685 kb being the longest contiguous stretch of DNA analyzed to date in humans. Another cluster of immune response genes, the lambda locus, responsible for the lambda light chain antibodies located on chromosome 22q11, has been recently mapped using contigs of YACs and cosmids (Fripiat *et al.*, 1995; Kawasaki *et al.*, 1995) and is the next candidate for total sequencing. This knowledge opens new perspectives to explore the normal and pathological function and evolution of the human immune system.

Publication supported by FAPESP.

REFERENCES

- Concannon, P., Pickering, L.A., Kung, P. and Hood, L. (1986). Diversity and structure of human T-cell receptor beta-chain variable region genes. *Proc. Natl. Acad. Sci. USA* 83: 6598-6602.
- Fripiat, J.-P., Williams, S.C., Tomlinson, I.M., Cook, G.P., Cherif, D., Le Paslier, D., Collins, J.E., Dunham, I., Winter, G. and Lefranc, M.P. (1995). Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2. *Hum. Mol. Genet.* 4: 983-991.
- Hesse, J.E., Lieber, M.R., Mizuuchi, K. and Gellert, M. (1989). V(D)J recombination: a functional definition of the joining signals. *Genes Dev.* 3: 1053-1061.
- Isobe, M., Erikson, J., Emmanuel, B.S., Nowell, P.C. and Croce, C.M. (1985). Location of the gene for β subunit of human T-cell receptor at band 7q35, a region prone to rearrangements in T-cells. *Science* 228: 580-582.
- Kabelitz, D. (1992). Function and specificity of human γ/δ -positive T-cells. *Crit. Rev. Immunol.* 11: 281-303.
- Kawasaki, K., Minoshima, S., Schooler, K., Kudoh, J., Asakawa, S., de Jong, P.J. and Shimizu, N. (1995). The organization of the human immunoglobulin λ gene locus. *Genome Res.* 5: 125-135.
- Kersh, G.J. and Allen, P.M. (1996). Essential flexibility in the T-cell recognition of antigen. *Nature* 380: 495-498.
- Kimura, N., Toyonaga, B., Yoshikai, Y., Du, R.P. and Mak, T.W. (1987). Sequences and repertoire of the human T-cell receptor alpha and beta chain variable region genes in thymocytes. *Eur. J. Immunol.* 17: 375-383.
- Koop, B.F., Rowen, L., Wang, K., Kuo, C.L., Seto, D., Lenstra, J.A., Howard, S., Shan, W., Deshpande, P. and Hood, L. (1994). The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region: organization, sequence, and evolution of 97.6 kb of DNA. *Genomics* 19: 478-493.
- Li, Y., Szabo, P. and Posnett, D.N. (1991). The genomic structure of human V beta 6 T-cell antigen receptor genes. *J. Exp. Med.* 174: 1537-1547. (Published erratum appears in *J. Exp. Med.* 1992 Vol. 175: 617).
- Mombaerts, P., Jacomini, J., Johnson, R.S., Herrup, K., Tonegawa, S. and Papaioannou, V.E. (1992). RAG-2 deficient mice have no mature B and T lymphocytes. *Cell* 68: 869-877.
- Oettinger, M.A., Schatz, D.G., Gorka, C. and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 248: 1517-1522.
- Raulet, D.H. (1989). Antigens for γ/δ T-cells. *Nature* 339: 342-343.
- Rowen, L., Koop, B.F. and Hood, L. (1996). The complete 685-kilobase DNA sequence of the human β T-cell receptor locus. *Science* 272: 1755-1762.
- Slightom, J.L., Siemieniak, D.R., Sieu, L.C., Koop, B.F. and Hood, L. (1994). Nucleotide sequence analysis of 77.7 kb of the human V beta T-cell receptor gene locus: direct primer-walking using cosmid template DNAs. *Genomics* 20: 149-168.
- Tillinghast, J.P., Behlke, M.A. and Loh, D.Y. (1986). Structure and diversity of the human T-cell receptor beta chain. *Science* 233: 879-883.
- Zhao, T.M., Whitaker, S.E. and Robinson, M.A. (1994). A genetically determined insertion/deletion related polymorphism in human T-cell receptor beta chain (TCRB) includes functional variable gene segments. *J. Exp. Med.* 180: 1405-1414.