

The Super-Information Age of Immunoglobulin Genetics

By Patrick C. Wilson and J. Donald Capra

From Molecular Immunogenetics, the Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73112

If the information age arose with the popularization of the personal computer, then the emergence of super-powered versions of these machines completely integrated into our lives and into the superstructure of society via the Internet has led us into the super-information age. Information is pouring into our lives and onto our desktops with unprecedented speed and volume. Genomics technology and the super-information age have led to a similar proliferation of information in the study of human genetics and immunobiology. In this issue of *The Journal of Experimental Medicine*, through innovation, persistence, and technological prowess, Matsuda et al. have completed sequencing the entire human V_H gene locus, including all of the immunoglobulin V_H genes as well as the hundreds of kilobases of intervening sequence (1). This landmark work and the efforts of several other groups of pioneering immunogeneticists in the study of the human V_H locus (2–5) have produced an extraordinary resource for the study of B cell biology and the important issues of immunoglobulin diversity. It appears that the study of immunoglobulin V_H genes in the human has also entered into the super-information era.

Tonegawa and the Information Age. With the final proof by Tonegawa (2) that the incredible diversity of immunoglobulin V gene sequences encoding the antibody repertoire were formed by the somatic recombination of relatively few genetic elements (for review see reference 6), the information age of V gene genetics was begun. In the almost twenty years since, a steady flux of one monumental discovery after another concerning the production and fine tuning of immunoglobulins and TCRs from their respective V genes has ensued. There have been extensive analyses of the molecular processes of V gene recombination and the factors involved, and of the intriguing but elusive process of somatic hypermutation. Analyses of the expressed V gene repertoire in the normal healthy subject and in association with immune defects and autoimmune syndromes have yielded steady and fruitful results as well. Although the V_H gene locus provided the framework for many of these important analyses, until now there has not been a complete sequence of this complex locus.

The Mapping Effort. In 1979, it was found through analyses of somatic cell hybrids between human lymphocytes to mouse myeloma cells that chromosome 14 was the only human chromosome present in all independent hybrids producing immunoglobulin heavy chains (3). Following this work, the quest to map and sequence the human V_H locus was underway. In the early 1980's, the V_H locus was further

mapped to chromosome band 14q32.33 (7, 8). The Honjo lab (whose work is highlighted in this commentary), along with several other key groups, was instrumental in mapping and deciphering as well as sequencing the human V_H locus. By the mid-1980's, Honjo's group began to piece together the V_H locus through Southern blot hybridizations to a library of cosmid clones encompassing 61 V_H genes (9), and, using progressively more advanced genome analysis techniques such as pulsed field gel electrophoresis and two-dimensional electrophoresis, several groups helped to frame the V_H locus by the beginning of this decade (10–12). With the addition of a second haplotype from isolated YAC clones, the Honjo lab constructed a map of the 3' end of the V_H locus in 1991 (13), which was extended in 1993 to 0.8 megabase encompassing the 64 most 3' V_H segments (4). The use of genomic Southern blot hybridization analyses provided an estimate at the time of the total complement of V_H genes, determined to be between 60 and 200 genes (10, 14). During this period, the expressed V_H gene repertoire was being compiled and appreciated as many of the V_H genes recognized today were cloned from a number of sources. With these advances, it was recognized that the V_H genes consist of seven families based on homology of >80% among V_H genes within each family (9, 10, 15–18). In 1992, in a significant contribution to this effort, Tomlinson et al. used a PCR-based approach using various sets of V gene family-specific primers to sequence 74 V genes designated as DP-1 through DP-74 from a single individual (17). These sequences became instrumental in the efforts of this group to map the V_H locus (see discussion of Cook and colleagues, below), and provided a good approximation of the total V_H gene complement of a single individual. The ongoing mapping efforts demonstrated that there is no significant clustering of the individual V_H gene families on the locus because they were found intermixed throughout. It was also recognized, primarily through cDNA sequence analysis of numerous autoantibodies performed in laboratories around the world, that although there is certainly allelism involving the individual human V_H genes (19–21) compared with other multigene families such as the HLA complex and to the mouse V_H genes, the human V_H locus exhibits relatively little polymorphism (22–24). This trait was an important factor in the efforts to sequence the human locus. However, it should be realized that there are differences between people in the presence or absence of single or blocks of V_H genes due to insertional/deletional polymorphism of the V_H locus (11, 13, 25).

Finally, in 1994, Cook and colleagues completed the map of the human V_H locus with the analysis of a second haplotype at the 3' end (25) and through an extension to the telomeric end of chromosome 14q32.3 (5). The final map of the V_H locus at this time was $\sim 1,100$ kb and included an estimated 95 V_H genes varying depending on the haplotype, with ~ 51 functional and the remainder pseudogenes. In addition, 24 orphan V_H genes not believed to contribute to the production of functional antibodies were found on chromosomes 15 and 16 (26–28).

About a Million Basepairs. In this issue of *The Journal of Experimental Medicine*, Matsuda et al. report having independently mapped the telomeric end of the V_H locus, and have completed sequencing the entire span of 957,090 bp (1). In their analysis, the V_H locus contains 123 V_H genes, including 39 functional genes known to produce heavy chains, 5 genes that appear functional but have not been reported as heavy chain proteins, and 79 pseudogenes. It is striking that approximately two-thirds of the V_H genes in the human locus are nonfunctional. Again it should be appreciated that the human V_H locus can contain insertional/deletional polymorphism depending on the particular haplotype (11, 13, 25). Previous estimates report that approximately half of the V_H genes are functional, and we have found transcripts for 10–12 V_H4 family genes in analyses of tonsils from five different individuals (Wilson, P.C., Y.J. Liu, J. Banchereau, V. Pascual, and J.D. Capra, unpublished results), compared with the seven transcribed V_H4 genes reported by Matsuda et al. in this issue (1). The extent and importance of V_H gene allelism between particular V genes or the total complement of V_H genes in different individuals, disease states, or racial groups is an area of interest that should not be superseded. Of particular interest concerning the differential complexity of V gene loci are the recent surprising findings of Green and Jakobovits that V gene complexity not only affects diversity, but is also important for the efficient development of the B cell lineage (29). With such observations in mind, it is clear that the sequence of a single or even a few human haplotypes does not tell the entire story, but it does tell the majority of the story and provides a vital framework for future analyses.

Of Man Not Mice. The idea that the human V_H locus would be completely mapped, let alone sequenced, long before the murine system would have been preposterous to

many immunogeneticists in the early 1980s. The human genome initiative provided both funds and intellectual cover for this huge amount of work. Additional impetus was directed with the association of various V_H genes with diseases from autoimmune syndromes to various lymphoid leukemias and lymphomas. Compounding these practical issues was the unforeseen and not yet fully appreciated complexity of the murine V_H locus relative to humans. In the mouse, there is extensive polymorphism both of the V gene locus in terms of the V gene complement of a particular haplotype and also between the individual genes. Thus, despite the considerably greater ease of genetic manipulation and analysis in the mouse, sequencing of the murine locus is not nearly as complete.

What the Future Holds. The super-information age of genomic analysis means that millions of bases of genetic code are being converted to bytes and the tools to analyze this information is increasingly being found on the desktop, not at the bench. In analyzing the development and expression of human V genes and somatic mutation, the bench work is increasingly just a matter of course and the real excitement comes only after stepping into the office. Having “our-favorite-locus” in its entirety only a URL away is a provoking thought and the list of analyses and old questions to be addressed using this resource is long and fruitful. With the entire one million nucleotides deposited in the database, immunologists and geneticists from around the world have a unique resource at their fingertips. This most complex of loci—unique in containing so many pseudogenes and gene fragments, and a locus that is (other than other V gene loci) unique in being so dynamic with gene segment recombination events—can now be studied in even more detail. With the single caveat of allelic polymorphism, we now “know what’s in the germline.” There can no longer be issues concerning the number of V, D, and J gene segments. Indeed, the great debate of germline versus somatic is now fully laid to rest. Another generation of scientists is looking at promoters, enhancers, tissue-specific factors, chromosomal end points, and the like. All of this is now available thanks to this pioneering work by Matsuda et al. (1). The study of human V_H genes has now entered the postgenomics era in which all human bioscience will be propelled in the near future via the human genome project.

Address correspondence to J. Donald Capra, Molecular Immunogenetics, The Oklahoma Medical Research Foundation, 825 NE 13th St., Oklahoma City, OK 73112. Phone: 405-271-7393; Fax: 405-271-8237; E-mail: the_capras@email.msn.com

Received for publication 30 September 1998.

References

1. Matsuda, F., K. Ishii, P. Bourvagnet, K.-i. Kuma, H. Haya-shida, T. Miyata, and T. Honjo. 1998. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J. Exp. Med.* 188:2151–2162.
2. Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature.* 302:575–581.
3. Croce, C.M., M. Shander, J. Martinis, L. Cicurel, G.G. D’Ancona, T.W. Dolby, and H. Koprowski. 1979. Chromo-

- somal location of the genes for human immunoglobulin heavy chains. *Proc. Natl. Acad. Sci. USA*. 76:3416–3419.
4. Matsuda, F., E.K. Shin, H. Nagaoka, R. Matsumura, M. Haino, Y. Fukita, S. Taka-ishi, T. Imai, J.H. Riley, R. Anand, and T. Honjo. 1993. Structure and physical map of 64 variable segments in the 3'0.8-megabase region of the human immunoglobulin heavy-chain locus. *Nat. Genet.* 3:88–94.
 5. Cook, G.P., I.M. Tomlinson, G. Walter, H. Riethman, N.P. Carter, L. Buluwela, G. Winter, and T.H. Rabbitts. 1994. A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. *Nat. Genet.* 7:162–168.
 6. Kindt, T.J., and J.D. Capra. 1984. *The Antibody Enigma*. Plenum Press, New York.
 7. Cox, D.W., V.D. Markovic, and I.E. Teshima. 1982. Genes for immunoglobulin heavy chains and for alpha 1-antitrypsin are localized to specific regions of chromosome 14q. *Nature*. 297:428–430.
 8. Kirsch, I.R., C.C. Morton, K. Nakahara, and P. Leder. 1982. Human immunoglobulin heavy chain genes map to a region of translocations in malignant B lymphocytes. *Science*. 216:301–303.
 9. Kodaira, M., T. Kinashi, I. Umemura, F. Matsuda, T. Noma, Y. Ono, and T. Honjo. 1986. Organization and evolution of variable region genes of the human immunoglobulin heavy chain. *J. Mol. Biol.* 190:529–541.
 10. Berman, J.E., S.J. Mellis, R. Pollock, C.L. Smith, H. Suh, B. Heinke, C. Kowal, U. Surti, L. Chess, and C.R. Cantor. 1988. Content and organization of the human Ig VH locus: definition of three new VH families and linkage to the Ig CH locus. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:727–738.
 11. Walter, M.A., U. Surti, M.H. Hofker, and D.W. Cox. 1990. The physical organization of the human immunoglobulin heavy chain gene complex. *EMBO (Eur. Mol. Biol. Organ.) J.* 9:3303–3313.
 12. van Dijk, K.W., L.A. Milner, and E.C. Milner. 1992. Mapping of human H chain V region genes (VH4) using deletion analysis and pulsed field gel electrophoresis. *J. Immunol.* 148:2923–2931.
 13. Shin, E.K., F. Matsuda, H. Nagaoka, Y. Fukita, T. Imai, K. Yokoyama, E. Soeda, and T. Honjo. 1991. Physical map of the 3' region of the human immunoglobulin heavy chain locus: clustering of autoantibody-related variable segments in one haplotype. *EMBO (Eur. Mol. Biol. Organ.) J.* 10:3641–3645.
 14. Matsuda, F., K.H. Lee, S. Nakai, T. Sato, M. Kodaira, S.Q. Zong, H. Ohno, S. Fukuhara, and T. Honjo. 1988. Dispersed localization of D segments in the human immunoglobulin heavy-chain locus. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1047–1051.
 15. Lee, K.H., F. Matsuda, T. Kinashi, M. Kodaira, and T. Honjo. 1987. A novel family of variable region genes of the human immunoglobulin heavy chain. *J. Mol. Biol.* 195:761–768.
 16. Shen, A., C. Humphries, P. Tucker, and F. Blattner. 1987. Human heavy-chain variable region gene family nonrandomly rearranged in familial chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA*. 84:8563–8567.
 17. Tomlinson, I.M., G. Walter, J.D. Marks, M.B. Llewelyn, and G. Winter. 1992. The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. *J. Mol. Biol.* 227:776–798.
 18. van Dijk, K.W., F. Mortari, P.M. Kirkham, H.W.J. Schroeder, and E.C. Milner. 1993. The human immunoglobulin VH7 gene family consists of a small, polymorphic group of six to eight gene segments dispersed throughout the VH locus. *Eur. J. Immunol.* 23:832–839.
 19. van Dijk, K.W., E.H. Sasso, and E.C. Milner. 1991. Polymorphism of the human Ig VH4 gene family. *J. Immunol.* 146:3646–3651.
 20. Sasso, E.H., K. Willems van Dijk, A.P. Bull, and E.C. Milner. 1993. A fetally expressed immunoglobulin VH1 gene belongs to a complex set of alleles. *J. Clin. Invest.* 91:2358–2367.
 21. Sasso, E.H., J.H. Buckner, and L.A. Suzuki. 1995. Ethnic differences of polymorphism of an immunoglobulin VH3 gene. *J. Clin. Invest.* 96:1591–1600.
 22. Chen, P.P., M.F. Liu, S. Sinha, and D.A. Carson. 1988. A 16/6 idiotype-positive anti-DNA antibody is encoded by a conserved VH gene with no somatic mutation. *Arthritis Rheum.* 31:1429–1431.
 23. Dersimonian, H., K.P. McAdam, C. Mackworth-Young, and B.D. Stollar. 1989. The recurrent expression of variable region segments in human IgM anti-DNA autoantibodies. *J. Immunol.* 142:4027–4033.
 24. Sanz, I., P. Kelly, C. Williams, S. Scholl, P. Tucker, and J.D. Capra. 1989. The smaller human VH gene families display remarkably little polymorphism. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:3741–3748.
 25. Walter, G., I.M. Tomlinson, G.P. Cook, G. Winter, T.H. Rabbitts, and P.H. Dear. 1993. HAPPY mapping of a YAC reveals alternative haplotypes in the human immunoglobulin VH locus. *Nucleic Acids Res.* 21:4524–4529.
 26. Matsuda, F., E.K. Shin, Y. Hirabayashi, H. Nagaoka, M.C. Yoshida, S.Q. Zong, and T. Honjo. 1990. Organization of variable region segments of the human immunoglobulin heavy chain: duplication of the D5 cluster within the locus and interchromosomal translocation of variable region segments. *EMBO (Eur. Mol. Biol. Organ.) J.* 9:2501–2506.
 27. Nagaoka, H., K. Ozawa, F. Matsuda, H. Hayashida, R. Matsumura, M. Haino, E.K. Shin, Y. Fukita, T. Imai, and R. Anand. 1994. Recent translocation of variable and diversity segments of the human immunoglobulin heavy chain from chromosome 14 to chromosomes 15 and 16. *Genomics*. 22:189–197.
 28. Tomlinson, I.M., G.P. Cook, N.P. Carter, R. Elasarapu, S. Smith, G. Walter, L. Buluwela, T.H. Rabbitts, and G. Winter. 1994. Human immunoglobulin VH and D segments on chromosomes 15q11.2 and 16p11.2. *Hum. Mol. Genet.* 3:853–860.
 29. Green, L.L., and A. Jakobovits. 1998. Regulation of B cell development by variable gene complexity in mice reconstituted with human immunoglobulin yeast artificial chromosomes. *J. Exp. Med.* 188:483–495.