

Minireview

The Immunoglobulin κ Gene Families of Human and Mouse: a Cottage Industry Approach*

Hans G. Zachau

Adolf-Butenandt-Institut der Universität München,
Molekularbiologie, Schillerstr. 44, D-80336 München,
Germany

Some aspects of the work of our group on the human and mouse immunoglobulin κ genes are reviewed. The human κ locus contains a large duplication: a 600 kb C κ -proximal copy with 40 V κ genes is found in the close vicinity of a 440 kb C κ -distal copy with 36 very similar, but not identical, V κ genes. The chimpanzee has only the C κ -proximal copy of the locus. The κ locus of the mouse is close to 3.2 Mb in size, of which 3.1 Mb have been cloned in four contigs, leaving three small gaps of together about 90 kb; 140 V κ genes and pseudogenes were localized and sequenced. In parallel to the elucidation of the structure of the κ loci, the mechanisms of the V-J rearrangement, somatic hypermutation and κ gene expression were studied. Various polymorphisms were detected in the human population and a number of haplotypes defined. In addition to the V κ genes within the loci numerous V κ orphans were localized on different chromosomes. Comparing the κ loci of different species allows some interesting conclusions as to the evolution of this multigene family. Finally our strategy of elucidating the structure and function of the κ loci, which has been termed a 'cottage industry approach', is discussed in relation to the large-scale genome analysis as pursued today using automated methods.

Key words: Contigs / Human / Immunoglobulin κ locus / Mouse.

The genes and loci of the immunoglobulin heavy and light chains have been reviewed by Honjo and Alt (1995) and by Max (1999). A wealth of structural and functional data is also available on the Internet (Immunoglobulin gene databases, 1999).

Work on Immunoglobulin Genes in Munich

In our laboratory the work on immunoglobulin genes started in the late '70s with experiments in which the chromatin structure of expressed κ genes in mouse myelomas was compared to the chromatin structure of the corresponding silent genes in mouse liver. In the following years both chromatin and immunoglobulin gene projects were pursued in our laboratory, but it turned out that the chromatin work was slow-moving and not very informative. On the other hand, at that time Tonegawa discovered the recombination of variable, joining and constant (V, J, C) gene segments and many researchers, including us, were intrigued by the new possibilities of cloning and analyzing mammalian genes. We found the reciprocal recombination products, now called signal joints (Steinmetz *et al.*, 1980), contributed to the understanding of somatic mutations (Pech *et al.*, 1981) and defined the immunoglobulin gene promoter (Falkner and Zachau, 1984). The last chromatin project in our lab came to an end in 1984. A few years later chromatin research picked up momentum worldwide again, but by then we were fully engaged in work on immunoglobulin genes and did not return to chromatin.

After working on immunoglobulin genes for a few years we felt that in order to understand certain functional and mechanistic aspects, we should know all V, J and C genes in at least one of the loci. Therefore in 1982 we embarked on an attempt to clone all human immunoglobulin V κ genes. 12 years later, when the elucidation of the human κ locus was almost complete, we started work on the mouse κ locus in parallel, the mouse being the main experimental animal in immunology. This project is now nearing completion. While the methods of genomic analysis still had to be developed during the work on the human κ locus, the advent of yeast and bacterial chromosomes (YACs and BACs) and of PCR, including long-range PCR in the early '90s, made the work on the mouse κ locus much easier.

Work on both the human and the mouse κ locus started with the screening of genomic libraries with V κ gene probes. The resulting clones were mapped in detail by restriction nuclease cleavage. Contigs (contiguously cloned regions) were constructed and the gaps closed by all tricks of chromosomal walking. We invested much effort in attempts to isolate all V κ genes. In parallel the restriction maps of the loci were established by pulsed field gel electrophoresis of genomic DNA, YAC and BAC digests. This type of structure elucidation is something like solving a giant puzzle.

* Published on the occasion of Hans G. Zachau's lecture at the Fall Meeting of the German Society for Biochemistry and Molecular Biology (GBM), 2000.

The human κ locus contains a large duplication: a 600 kb C κ -proximal copy with 40 V κ genes is found near a 440 kb C κ -distal copy with 36 very similar, but not identical, V κ genes. We sequenced only the gene regions, recombination breakpoints and other areas of special interest, but we never engaged in automated large scale sequencing. However, we gave cosmid and phage λ clones covering the whole locus to N. Shimizu and K. Kawasaki in Tokyo, who are sequencing them.

Most V κ genes in the proximal copy of the human κ locus rearrange to JC by a deletion mechanism while, in an act of 'genome acrobatics', the distal ones use an inversion mechanism (Weichhold *et al.*, 1990). Somatic hypermutation is of continuing interest to us (*e.g.* Klein *et al.*, 1993). In studying polymorphisms in human populations we have defined a number of haplotypes. In one haplotype (No. 11) the whole distal copy of the locus is deleted (Pargent *et al.*, 1991). With the help of an individual who is homozygous for this deletion we could study the function of a specific distal gene (A2), which codes for the most common light chain in the *Haemophilus influenzae* response. This in turn offered the possibility to do some interesting vaccination experiments (Scott *et al.*, 1992). The κ locus is located on the short arm of chromosome 2, but 25 so-called V κ orphans were found on different chromosomes and also on the long arm of chromosome 2 (Lötscher *et al.*, 1986; Zachau, 1995). The work on the human V κ genes has been reviewed earlier (Zachau, 1995, 1996).

The κ locus of the mouse (C57BL/6) is close to 3.2 Mb in size, of which 3.1 Mb have been cloned in two large and two small contigs. We were not able to close the last three small gaps of together about 90 kb before my retirement, but we hope this will soon be done by others. 140 V κ genes and pseudogenes, as well as a few V κ relics and orphans, were localized and sequenced (Kirschbaum *et al.*, 1999; Thiebe *et al.*, 1999; Rösenthaller *et al.*, 1999, 2000; a detailed restriction map of the locus and supplementary information can be found on our homepage). The mouse κ locus is three times larger and has nearly twice as many V κ genes as the human locus. In the mouse locus there is only one fairly small internal duplication. The V κ genes belong to 18 distinct gene families and are arranged in different 5', 3' polarities.

Both the human and the mouse κ locus projects have been pleasant or at least satisfying ventures for me and for most members of the group. A few senior colleagues, several German and foreign postdoctoral fellows and many doctoral students rendered their contributions. During most of the time there were an average of 10 scientists (including doctoral students) in the group who were supported by a number of technicians. 34 doctoral theses and several diploma theses developed from our work on the immunoglobulin genes. In most of the theses structural work was combined with mechanistic or functional studies. Our experiments with the human κ genes were described in 70 publications, the work with the mouse genes in 34 papers, some also in this journal and many of the recent ones in the European Journal of Immunology.

What About the Evolution of the κ Locus?

In the human κ locus the sequence of the gene regions of the two copies differ on the average by 1%, indicating that the duplication may have taken place a million years ago. This is a soft number for various reasons, but it is roughly in keeping with the finding that chimpanzees have just the proximal half of the locus (Ermer *et al.*, 1995). Apparently the duplication took place after the human and chimp clades diverged, which is assumed to have occurred 5.5 million years ago (Kumar and Hedges, 1998).

Although some human and chimp genes are very similar (the C κ genes are 99.6% identical), one should not overemphasize the similarity between the genomes of the two species. Not meaningful at the present state of knowledge are questions of the type: 'where are the 1 or 2% of sequence differences located in the genome that are responsible for the differences between human and chimp?' The duplication of the κ locus may serve as an example of the type of differences that could be of functional importance if they occur for instance in a region of the genome that is important for brain function. Duplications of amplifications of genomic regions cannot be easily detected by overall hybridization experiments or by the current comparisons of scattered gene sequences; they will become known (and hopefully interpretable) once both genomes have been fully sequenced.

Many events occurred (long) before the duplication of the human κ locus, *e.g.* the amplification and interdigitation of V κ genes of different subgroups, most changes which converted functional genes to pseudogenes, the transposition of V κ orphans to other chromosomes and (by a pericentric inversion) to the long arm of chromosome 2. The latter process apparently happened after the gorilla, and before the chimp, clades diverged from the human evolutionary tree (Arnold *et al.*, 1995). Pericentric inversions still occur in the human population; 0.1% of the population have the κ locus on the long arm of chromosome 2 and the orphans on its short arm (Lautner-Rieske *et al.*, 1993).

Events that occurred after the duplication of the human κ locus include, for instance, small deletions found only in one copy of the κ locus, insertion of an Alu element into one but not the other copy, or numerous point mutations (twice as many transitions than transversions), which are the main feature of the observed divergence of the two copies of the locus.

The fairly large difference between the human and the mouse κ loci is not surprising considering the fact that the two clades diverged on a molecular timescale about 40 million years ago (Kumar and Hedges, 1998). Still, some related genes occur in homologous positions in the mouse and human κ loci, and there are sequence similarities of 74 – 82.5% between several gene families of the two species. (For comparison: within one species genes with $\geq 80\%$ sequence identity by definition belong to the same family).

The Cottage Industry Approach, Genomics and Postgenomics

Our work on the immunoglobulin κ genes was aimed at the elucidation of the biologically interesting structures of the loci, *i.e.* the coding regions, regulatory sequences and other sites of functional importance. This type of work used to be classified as molecular biology or genomics. In the past 10 years laboratory robots and sequencing machines have been developed, making large-scale or even huge-scale restriction mapping and sequencing possible. In order to interpret the automatically generated sequences, subsequent functional and mechanistic studies, termed 'postgenomics', are necessary. Our strategy to combine, in a laboratory without big machines, structural work and functional studies was termed 'the cottage industry approach' by one of the moguls of the automated genomics era.

Obviously the cottage industry approach was the only one possible for much of the 20 years in which it was practiced in our laboratory, and the approach paid heavy dividends. Our data on the human κ locus became available about 10 years before the sequence of the whole genome is scheduled to be completed. The situation with respect to the mouse κ locus may turn out to be similar. This means that several groups, including ours, could make use of the data for mechanistic and functional studies, some of which are mentioned in this minireview. We believe that many interesting questions concerning the generation of antibody diversity, for instance, have been answered already. In other words, for the κ loci much of the so-called postgenomic work has been done, or is currently being done.

What new insights will be gained when the κ loci of human and mouse are sequenced completely during the course of the current large-scale sequencing effort? The main point, as far as I can see, will be that the knowledge of the sequences between the $V\kappa$ genes and beyond the borders of the loci will allow a more detailed understanding of genome reshuffling and evolution. Also, one or the other $V\kappa$ gene and several relics of $V\kappa$ genes may be found that we missed in our largely hybridization-based approach.

The present genomics revolution is extremely important. Many techniques and experiences developed in the cottage industry are of course now incorporated into the automated approaches. Nobody would apply today the classical molecular biology techniques without automation to the elucidation of a multigene family located in a few Mb of the human genome. Once all the data of the human genome are available in the computer, one will define the sequences of interest and perform the necessary functional tests. This way one will get the results much easier and in a fraction of the time required in the past.

Acknowledgements

I thank the members of our group for their dedicated work and for contributing to the friendly atmosphere in the laboratory. The work was supported in the beginning by Deutsche Forschungsgemeinschaft and later by Bundesministerium für Forschung und Technologie, which then became Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie; the ministries acted *via* Genzentrum München and more recently *via* Deutsches Humangenomprojekt. Continuous support was provided by Fonds der Chemischen Industrie.

References

- Arnold, N., Wienberg, J., Ermert, K., and Zachau, H.G. (1995). Comparative mapping of DNA probes derived from the $V\kappa$ immunoglobulin gene regions on human and great ape chromosomes by fluorescence in situ hybridization. *Genomics* 26, 147 – 150.
- Ermert, K., Mitlöchner, H., Schempp, W., and Zachau, H.G. (1995). The immunoglobulin κ locus of primates. *Genomics* 25, 623 – 629.
- Falkner, F.G., and Zachau, H.G. (1984). Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. *Nature* 310, 71 – 74.
- Honjo, T., and Alt, F.W., eds. *Immunoglobulin Genes*, 2nd Edition (London, New York: Academic Press).
- Immunoglobulin gene databases (1999). ImMunoGeneTics database of M.-P. Lefranc: <http://imgt.cnusc.fr:8104> and of I.M. Tomlinson: <http://www.mrc.cpe-cam.ac.uk/imt-doc/public/INTRO.html>.
- Kirschbaum, T., Rösenthaller, F., Bensch, A., Hölscher, B., Lautner-Rieske, A., Ohnrich, M., Pourrajabi, S., Schwendinger, J., Zocher, I., and Zachau, H.G. (1999). The central part of the mouse immunoglobulin κ locus. *Eur. J. Immunol.* 29, 2057 – 2064.
- Klein, R., Jaenichen, R., and Zachau, H.G. (1993). Expressed human immunoglobulin κ genes and their hypermutation. *Eur. J. Immunol.* 23, 3248 – 3262.
- Kumar, S., and Hedges, S.B. (1998). A molecular timescale for vertebrate evolution. *Nature* 392, 917 – 920.
- Lautner-Rieske, A., Hameister, H., Barbi, G., and Zachau, H.G. (1993). Mapping immunoglobulin gene related DNA probes to the central region of normal and pericentrically inverted human chromosome 2. *Genomics* 16, 497 – 502.
- Lötscher, E., Grzeschik, K.-H., Bauer, H.-G., Pohlenz, H.-D., Straubinger, B., and Zachau, H.G. (1986). Dispersed human immunoglobulin κ light chain genes. *Nature* 320, 456 – 458.
- Max, E.E. (1999). Immunoglobulins: Molecular Genetics. In: *Fundamental Immunology*, 4th edition, W.E. Paul, ed., (New York, USA: Raven Press Ltd.), pp. 315 – 382.
- Pargent, W., Schäble, K.F., and Zachau, H.G. (1991). Polymorphisms and haplotypes in the human immunoglobulin κ locus. *Eur. J. Immunol.* 21, 1829 – 1835.
- Pech, M., Höchtel, J., Schnell, H., and Zachau, H.G. (1981). Differences between germline and rearranged immunoglobulin $V\kappa$ -coding sequences suggest a localized mutation mechanism. *Nature* 291, 668 – 670.
- Rösenthaller, F., Kirschbaum, T., Heim, V., Kirschbaum, V., Schäble, K.F., Schwendinger, J., Zocher I., and Zachau, H.G. (1999). The 5' part of the mouse immunoglobulin κ locus. *Eur. J. Immunol.* 29, 2065 – 2071.

- Röschenthaler, F., Hameister, H., and Zachau, H.G. (2000). The 5' part of the immunoglobulin κ locus as a continuously cloned structure. *Eur. J. Immunol.*, in press.
- Scott, M.G., Zachau, H.G., and Nahm, M.H. (1992). The human antibody V region repertoire to the type b capsular polysaccharide of *Haemophilus influenzae*. *Intern. Rev. Immunol.* 9, 45 – 55.
- Steinmetz, M., Altenburger, W., and Zachau, H.G. (1980). A rearranged DNA sequence possibly related to the translocation of immunoglobulin gene segments. *Nucleic Acids Res.* 8, 1709 – 1720.
- Thiebe, R., Schäble, K.F., Bensch, A., Brensing-Küppers, J., Heim, V., Kirschbaum, T., Mittlöhner, H., Ohnrich, M., Pourrajabi, S., Röschenthaler, F., Schwendinger, J., Wichelhaus, D., Zocher, I., and Zachau, H.G. (1999). The variable genes and gene families of the mouse immunoglobulin κ locus. *Eur. J. Immunol.* 29, 2072 – 2081.
- Weichhold, G.M., Klobeck, H.-G., Ohnheiser, R., Combriato, G., and Zachau, H.G. (1990). Megabase inversions in the human genome as physiological events. *Nature* 347, 90 – 92.
- Zachau, H.G. (1995). The human immunoglobulin κ genes. In: *Immunoglobulin Genes*, 2nd Edition, T. Honjo and F.W. Alt, eds., (London, New York: Academic Press) 173 – 191.
- Zachau, H.G. (1996). The human immunoglobulin κ genes. *The Immunologist* 4, 49 – 54.