Nucleotide sequence of the intron of the germline human κ immunoglobulin gene connecting the J and C regions reveals a matrix association region (MAR) next to the enhancer

Charles Whitehurst¹, Henry R. Henney^{1,2}, Edward E. Max³, Harry W. Schroeder Jr⁴, Frank Stüber⁴, Katherine A. Siminovitch⁵ and William T. Garrard^{1*}

¹Department of Biochemistry, UT Southwestern Medical Center, Dallas, TX 75235, ²Department of Biology, University of Houston, Houston, TX 77204, ³National Institutes of Health, Bethesda, MD 20892, ⁴Departments of Medicine and Microbiology, University of Alabama, Birmingham, AL 35294, USA and ⁵Department of Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada

Submitted July 22, 1992

EMBL accession no. X67858

DNA within interphase nuclei and mitotic chromosomes is believed to be organized into topologically constrained looped domains ranging from 5 to 200 kb in length (reviewed in refs. 1, 2). *In vitro* DNA binding assays have identified sequences termed MARs ('matrix association regions'), also called SARs ('scaffold attached regions'), that are thought to mediate loop attachment *in vivo* (3, 4). These sequences are at least 200 bp long, AT-rich (ca. 70%), and often contain topoisomerase II and other AT-rich consensus motifs (1). MARs appear to be evolutionarily conserved, since similar sequences have been identified in DNA derived from human, mouse, hamster, chicken, rabbit, *Drosophila* and yeast, as well as viral genomes. These elements also appear to have prokaryotic counterparts (1).

Particularly intriguing is the observation that the κ immunoglobulin genes of the mouse and rabbit have MARs adjacent to their intronic transcriptional enhancer elements (3, 5). Stable integration experiments with constructs lacking the mouse MAR and natural mutations of the corresponding region in the rabbit gene reveal that the MARs are necessary for maximal gene expression (6, 7). The mouse immunoglobulin heavy chain gene enhancer also is flanked by MARs (8), and a MAR has been identified within the chicken λ immunoglobulin constant region locus (9). Here we address the question of whether a MAR resides within the human κ immunoglobulin gene in a position similar to that of the mouse and rabbit counterparts.

Although portions of the human κ gene locus have been sequenced (10–14), the relevant region predicted to contain a MAR remained to be characterized (Figure 1). We therefore resequenced segments of the locus including an additional previously unreported 1.9 kb and thereby generated a contiguous updated sequence spanning 5364 nucleotides (Accession no. X67858 in the EMBL Database). Based on the AT-content, frequency of ATATTT, and topoisomerase II consensus sequences (3), we conclude that a MAR resides within the human κ gene just upstream of the enhancer (Figure 1). Dot matrix analysis further confirms this assignment since the human MAR shares sequence similarity with its mouse and rabbit counterparts (Figure 2).

ACKNOWLEDGEMENTS

This investigation was supported by grants GM22201, GM29935, and GM31689 from the National Institutes of Health and grant I-823 from the Robert A. Welch Foundation to WTG, by grant SG-192 from the American Cancer Society to HRH, by grants AI23694, AI30879, AR03555, CA45232, and AR20216 from the National Institutes of Health to HWS and by grant MT-10730 from the Medical Research Council of Canada to KAS. FS was supported by the Deutscher Akademischer Austauschdienst. HWS is an RJR Research Scholar in Immunology.

REFERENCES

- Garrard, W.T. (1990) In: Eckstein, F. and Lilley, D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin. Vol. 4, pp. 163–175.
- Freeman, L.A. and Garrard, W.T. (1992) in: Stein, G.S., Stein, J.L. and Lian, J.B. (eds), *Critical reviews in eukaryotic gene expression*. CRC Press, Inc. Vol. 2, pp. 165–209.

Figure 1. Structure and sequencing strategy of the human κ gene J-C region. Numbers 10–14 refer to the references for previously published sequences within the bracketed regions. a, sequence determined from cloned DNA by the Maxam–Gilbert technique; b, sequence determined from cloned DNA and/or PCR amplified DNA of several different individuals by the dideoxy technique. Indicated are the joining (J), MAR (M), enhancer (E), constant region (C), and above are matches to the topo2 consensus (▲), ATATTT (arrows), and sequences greater than 70% AT (bracket).

Figure 2. Dot matrix sequence comparisons between the human α gene J-C region and those of the mouse and rabbit genes. The central diagonal strip represents a comparison between rabbit and mouse genes. M represents the MAR. An additional 394 bp of upstream sequence (derived only from one direction of sequencing) used in this analysis is available from the authors.

3. Cockerill, P.N. and Garrard, W.T. (1986) *Cell* **44**, 273–282.
4. Gasser, S.M. and Laemmli, U.K. (1986) *Cell* **46**, 521–530.
5. Sperry, A.O., Blasquez, V.C. and Garrard, W.T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5497–5501.
6. Blasquez, V.C., Xu, M., Moses, S.C. and Garrard, W.T. (1989) *J. Biol. Chem.* **264**, 21183–21189.
7. Xu, M., Hammer, R.E., Blasquez, V.C., Jones, S.L. and Garrard, W.T. (1989) *J. Biol. Chem.* **264**, 21190–21195.
8. Cockerill, P.N., Yuen, M.-H. and Garrard, W.T. (1987) *J. Biol. Chem.* **262**, 5394–5397.
9. Parvari, R., Ziv, E., Lantner, F., Heller, D. and Schechter, I. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3072–3076.
10. Hieter, P.A., Max, E.E., Seidman, J.G., Maizel, J.V. Jr and Leder, P. (1980) *Cell* **22**, 197–207.
11. Hieter, P.A., Maizel, J.V. Jr and Leder, P. (1982) *J. Biol. Chem.* **257**, 1516–1522.
12. Emorine, L., Kuehl, M., Weir, L., Leder, P. and Max, E.E. (1983) *Nature* **304**, 447–449.
13. Klobeck, H.-G. and Zachau, H.G. (1986) *Nucleic Acids Res.* **14**, 4591–4603.
14. Kato, S., Tachibana, K., Takayama, N., Kataoka, H., Yoshida, M.C. and Takano, T. (1991) *Gene* **97**, 239–244.