

## GENOMIC CLONING OF THE HUMAN HISTO-BLOOD GROUP ABO LOCUS

Eric Paul Bennett<sup>1</sup>, Rudi Steffensen<sup>1,2</sup>, Henrik Clausen<sup>1\*</sup>, Daniël Olde Weghuis<sup>3</sup>  
and Ad Geurts van Kessel<sup>3</sup>

<sup>1</sup>Faculty of Health Sciences, School of Dentistry, University of Copenhagen, Copenhagen, Denmark

<sup>2</sup>Aalborg Blood Bank, Denmark

<sup>3</sup>University Hospital Nijmegen, Dept. of Human Genetics, The Netherlands

Received November 14, 1994

---

**SUMMARY:** A P1 phage clone containing the near complete protein coding region of the human histo-blood group ABO gene locus was isolated and characterized. The insert was estimated at  $\approx 41$  kp, and an  $\approx 20$  kbp *Bam*HI fragment contained the near complete coding region as evaluated by Southern blot analysis. Intron/exon boundaries were determined and the coding region was found to span at least seven exons. *In situ* hybridization to metaphase chromosomes confirmed the predicted localization to 9q34. © 1995 Academic Press, Inc.

---

The histo-blood group ABO system is one of the major allogeneic antigen systems in man. The chemistry, biosynthesis, and molecular genetics have been established (1-3). Recent work involving isolation of the blood group A glycosyltransferase and cDNA cloning of ABO alleles has confirmed that the A and B glycosyltransferases are alleles encoded by the ABO locus (4,5). The polymorphism involved in the ABO blood group system including that of rare subgroups has been found to represent minute structural changes in the coding region of the ABO alleles (for a recent review see (6)).

Here we present the genomic structure of the coding region of the ABO locus and confirm by *in situ* hybridization to metaphase chromosomes the localization to 9q34.

### MATERIAL AND METHODS

**Isolation and characterization of a genomic P1 phage clone:** A human P1 library (DuPont Merck Pharmaceutical Company Human Forskin Fibroblast P1 Library) was

---

\*Correspondence: School of Dentistry, Nørre Alle 20, DK-2200 Copenhagen N, Denmark.  
Fax nr.: +45 3532 6505.

screened using PCR primers FY57/FY46 (7) and a single positive clone DMPC-HFF#1-0529F4 was obtained (Custom Service Genome System Inc.) Plasmid DNA was prepared according to the supplier. The P1 DNA was partly sequenced by use of sense and antisense primers located in the coding region of the ABO locus. Double stranded sequencing was performed using a modified Perkin-Elmer cycle-sequencing kit on a model 480 thermocycler (Perkin-Elmer). Intron sizes were determined by PCR amplification using flanking primers followed by hybridization with specific probes to the intron products.

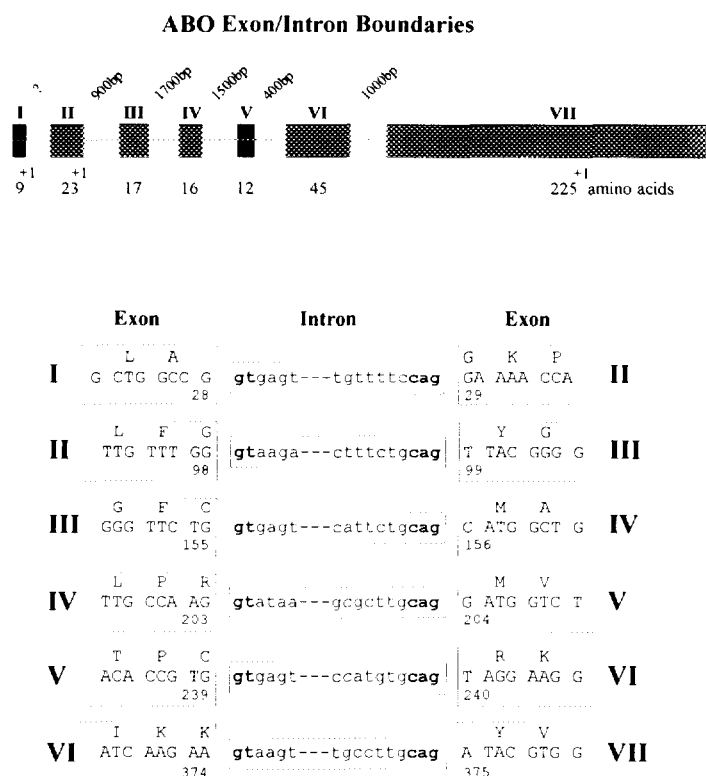
**Expression of A transferase in Sf9 cells:** The putative soluble A transferase gene was produced by RT-PCR of poly(A)<sup>+</sup> RNA from gastric cancer cell line MKN45 using the primers EPB83/EPB90 (5'-TCGAATTCGCCCCAGAAGTCTAATGCCAG and 5'-TCGAATTCTGGCAGCCGCTCACGGGTTCC) located in exon III and VII, respectively (7). The DNA was cloned into the EcoRI site of pT7T3U19 (Pharmacia), followed by subcloning into the EcoRI site of the baculovirus transfer vector pACGP67. Plasmid pACGP67-A-transferase was cotransfected with purified baculo-Gold DNA (PharMingen) and recombinant baculovirus was obtained by three successive amplifications in Sf9 cells according to PharMingen's protocol. Blood group A enzyme activity in the culture medium was assayed as previously described using 2-fucosyllactose as acceptor substrate and UDP-[<sup>14</sup>C]-GalNAc as donor substrate (4). Controls included noninfected cells and cells infected with a construct containing UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase T2 (unpublished).

**RT-PCR analysis of ABO gene transcripts in human organs:** RT-PCR was performed using primers EPB79/EPB88 (5'-ACGAATTCTACTTGTTTCAGGTGGCTCTCGTC and 5'-TCGAATTCATGGCCGAGGTGTTGCGGAC) located in exons I and VII respectively, of which EPB79 was used as RT primer (RT-PCR kit, Perkin-Elmer). PCR products were probed with an exon VII specific probe EPB77 (5'-CTCCGCTGTTCCGGCACC) (Fig. 3, panel A).

**In situ hybridization to metaphase chromosomes:** P1 DNA was labeled with biotin-14-dATP using the bio-NICK system (Life Technologies). The labeled DNA was precipitated with ethanol in the presence of herring sperm DNA. Precipitated DNA was dissolved and denatured at 80°C for 10 min followed by incubation for 30 min at 37°C and added to heat-denatured chromosome spreads where hybridization was carried out over night in a moist chamber at 37°C. After posthybridization washing (50% formamide, 2 x SSC at 42°C) and blocking with nonfat dry milk powder, the hybridized probe was detected with avidin-FITC (Vector Laboratories) followed by two amplification steps using rabbit-anti-FITC (Dako) and mouse-anti-rabbit FITC (Jackson Immunoresearch), chromosome spreads were mounted in antifade solution with blue dye DAPI.

## RESULTS AND DISCUSSION

**Structure of the ABO gene locus:** A human P1 library was screened using a PCR probe based on two primers located in the immediate intron boundaries to exon VI yielding a product of 240 bp. A single clone DMPC-HFF#1-0529F4 was isolated and Southern analysis with various oligonucleotides covering the 3' and 5' coding sequence of the existing full length ABO cDNA indicated that the entire coding sequence apart from the most 3' sequence of the known last exon VII was included in an ≈ 20 kbp BamHI fragment (not shown). The structure of the coding region of exon VII is inferred from PCR-analysis of genomic DNA (6).



**Figure 1.** *Genomic organization of human ABO locus.* Diagrammatic representation of exon organization of the protein coding sequences (shaded). All exon regions were located on a single *Bam*HI fragment of a P1 clone and approximate intron sizes estimated by PCR are indicated where possible. Indicated are the exon size as coding amino acid number. Indicated by superscripts to the exon size are bases deviating from a multimer of three.

A comparative Southern blot analysis between the cloned P1 DNA and total human genomic DNA using a full length cDNA as probe gave identical/similar patterns (not shown). The P1 clone was partially sequenced and sequences of the coding region and 50-150 bp of adjacent introns were obtained (Figs. 1 and 2). As only little information of the untranslated 5' sequence of the ABO transcripts are available, the exon structure outside the coding region is still to be determined. Furthermore the 5' and 3' borders of exon 1 and exon 7, respectively, are undetermined. All exon/intron boundaries identified conformed to the GT-AG consensus rule (8).

The entire coding region of the to-date identified ABO cDNA clones is covered by 7 exons where the two largest, VI and VII, encode 77% of the full coding region and 91% of the coding region of the previously established

5' - GGCTCCGCGTCTGGTCTCCGCCCTCCGCCCTCCGCCCGCGGCCCTGGGGC -  
- *Exon I* -  
- GTGAGTGCAGGCCCTCGGCCCCGGGTGCCCGCAGGAGCCGCTACCGCAGG - 3'

5' - GGTGAGAGAAGGAGGGTGAGTGATGTGATTTTTCTACTCCTGTTTTCCAG -  
- *Exon II* -  
- GTAAGACACATTTGACCATCGAGGCTGGCCTGGTTTGGGGAGAAGTGACC - 3'

5' - AGGCAGTCTTCGTCTCTGTGCTCTGAGTCTTGGCACACTTCTCTTCTGCAG -  
- *Exon III* -  
- GTGAGTGCAGGGAAGAGCAGGTGGAGCATCCATGCTGGCCGGGGTCTGG - 3'

5' - TAAATCCTGCTCCTAGACTAACTTCATCTCCTGTGTTCTCATTCTGCAG -  
- *Exon IV* -  
- GTATAATGTCAAGTGCCTCCCTTCAGTGGCTCCCATGTACAGAATTGTCC - 3'

5' - TGCATCCACGCTTTCCATGCAAACCTCACGTGGCGCGCTTTGCGCTTGCAG -  
- *Exon V* -  
- GTGAGTAAAGTTACTGACACTGAAACTGAACGCAGCTCAAGGGGCTGTTC - 3'

5' - GGTGGCACCCCTGCCAGCTCCATGTGACCGCACGCGCTCTCTCCATGTGCAG -  
- *Exon VI* -  
- GTAAGTCAGTGAGGTGGCCGAGGGTAGAGACCCAGGCAGTGGCGAGTGAC - 3'

5' - AGGGGTGCACGGCCGGCGGCTCCCCAGCCCCGCTCCGCTGCCTTGCAG -  
- *Exon VII* -  
- GCGGCTGCCAGGGGCTCTGGGAGGGCTGCCGGCAGCCCCGTCCCCCTCCC - 3'

**Figure 2.** Flanking untranslated sequences of exon I and VII and flanking intron sequences of exon II-VI.

catalytically active, soluble transferase protein. As shown in Table I, a PCR generated cDNA clone initiated at nucleotide position 115 (amino acid position 39) in exon III and including the entire 3' coding sequence resulted in a catalytically active protein when expressed in the baculovirus system. This is in good agreement with the observed amino-terminal sequence of purified soluble transferase protein from human lung (4) and intestinal mucosa (9).

Table I. *Expression of soluble A transferase in Sf9 cells*

Construct	Specific activity units*/ml
Uninfected cells	0.00005
A enzyme sol.-1	0.0054
A enzyme sol.-2	0.0050
GalNAc-T2 sol.1	0.00010

\* One unit defined as amount of enzyme transferring 1  $\mu$ mol GalNAc from UDP-GalNAc in 1 min using 2-fucosyllactose as acceptor.

Studies of the histo-blood group ABO polymorphism, especially of rare subgroups, have generally been limited to the coding region covered by exon VI and VII, as these could be analysed by PCR amplification of genomic DNA (6). The difficulties in obtaining RNA for these studies make it valuable to be able to PCR amplify each of the coding exons separately using primers located in adjacent introns for subsequent sequence analysis. Figure 2 presents the immediate 50 bp 5' and 3' intron sequences adjacent to exons, as well as 50 bp untranslated sequences 5' to exon I and 3' to exon VII. Studies of the molecular genetic basis of the blood group ABO subgroups have indicated that structural changes in exon VI and VII result in the observed blood group phenotype. However, limiting the analysis to exon VI and VII of the rare blood group subgroups A<sub>3</sub> and B<sub>3</sub> failed to provide a basis for these phenotypes and it is thus likely that analysis of the 5' end will show changes leading to the observed phenotype (10). Although the catalytic domain of the A/B glycosyltransferases appears mainly to be encoded by exon VI and VII, it is possible that changes more 5' in the gene may result in expression failure or changes in catalytic activity.

**Comparison of genomic structure of ABO and  $\alpha$ 1,3-galactosyltransferase:** The human ABO genes show considerable homology with the bovine/murine  $\alpha$ 1,3-galactosyltransferase as well as human processed pseudogenes hereof (11-13). As pointed out by Joziassse (14) the homology is limited to the last two coding exons in both the ABO gene (exon VI and VII) and the murine  $\alpha$ 1,3-galactosyltransferase gene (exon VIII and IX) and the sequence homology as well as same exon sizes suggest an evolutionary relationship based on exon-shuffling. The intron/exon structure in the remaining 5' part of the ABO locus shows similarity to that of the murine where 5, respectively 4, small exons code for the amino-terminal sequence. Interestingly, a similar structure in the reading frame of exons is also found. The ABO gene has a single base excess in the first two small exons (I and II), and the intervening exons III-VI are perfect multimers of three, whereas the last coding exon has a single base excess (Fig. 1). The bovine  $\alpha$ 1,3-galactosyl transferase shows the same pattern only deviating by having one coding exon less and thus two excess bases in the first coding exon (exon IV) (11). Alternative splicing of the ABO genes may thus exclude exons III-VI without changes in reading frame. RT-PCR using primers located in exon I and VII gave rise to varying product sizes in different human organs, thus indicating that transcripts with alternative exon usage are found in different organs (Fig. 3). Subcloning and sequencing of several of these PCR products verified the transcript structures indicated in Fig. 3. To what extent such

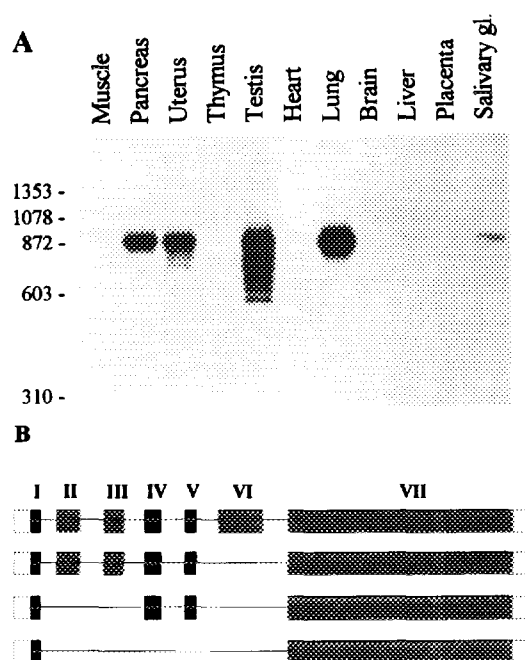
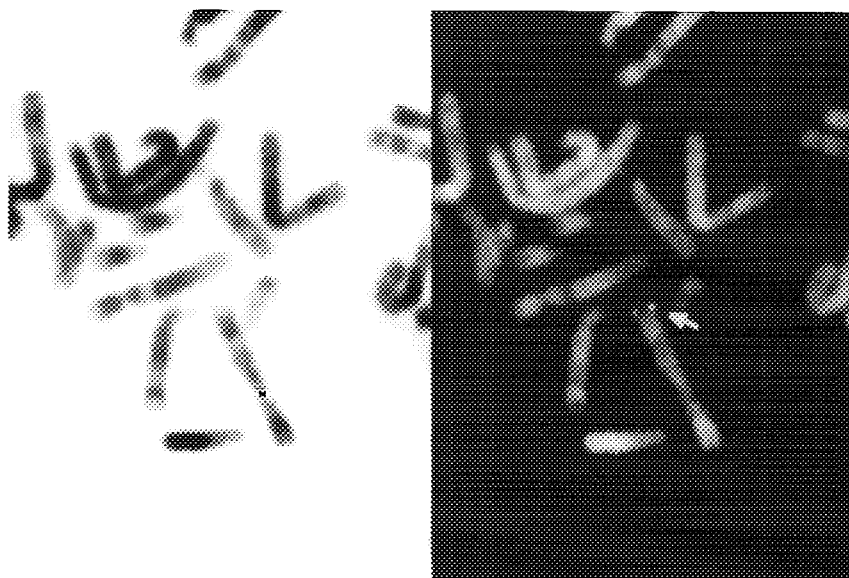


Figure 3. RT-PCR analysis of ABO gene transcripts in human organs. Panel A: RT-PCR products probed with exon VII specific probe. Panel B: Schematic representation of RT-PCR analysis of testis transcripts.

alternative transcripts are translated and enzymatically active will be the focus of future research. The finding of transcripts missing only exon VI is interesting as this contains the blood group O<sup>1</sup> frameshift mutation. A transcript lacking exclusively exon VI will introduce a stop codon and thus an enzymatically inactive truncated protein. Interestingly, Joziassse et al. (11) found no transcripts lacking this exon with the murine  $\alpha$ 1,3-galactosyltransferase. It seems clear from studies of truncated forms of the  $\alpha$ 1,3-galactosyltransferase that the protein sequence of both exon 8 and 9 corresponding to VI and VII in the ABO gene is necessary for transferase activity (15).

**Chromosomal localization of ABO locus:** Fluorescent in situ hybridization to metaphase chromosomes using the isolated P1 phage showed fluorescence signals at 9q34 (Fig. 4; 20 metaphases evaluated). No specific hybridization was observed at any other chromosomal site. This is in agreement with reported linkage analysis studies (16). Most of the glycosyltransferase genes for which a chromosomal localization has been established have shown that these are distributed all over the genome (14,17). Two processed pseudo-genes showing homology to  $\alpha$ 1-3-



**Figure 4.** *Localization of the DMPC-HFF#1-0529F4 probe on 9q34 using fluorescence in situ hybridization (right panel, arrow). The identification of individual chromosomes was deduced from converted DAPI banding patterns (left panel).*

galactosyltransferase and ABO gene are located on chromosome 12q14-q15 and 9q33-q34 (15), thus at least one co-localizes with the ABO locus.

The structure of the ABO locus reported here should enable more detailed studies on the polymorphism involved in rare ABO blood groups. The reported intron sequences will allow PCR strategies for complete analysis of the coding region of the ABO locus using readily available genomic DNA. The isolation of a P1 clone covering at least the entire 5' coding region of the ABO locus will enable studies of the promoter region.

**Acknowledgments.** A similar structure of the coding region of the ABO gene has been obtained using multiple overlapping DNA clones (F. Yamamoto and S. Hakomori personal communication). This work was supported by the Danish Cancer Society, The Lundbeck Foundation, The Danish Medical Council, and the Foundation of Voluntary Blood Donors.

#### REFERENCES

1. Watkins, W.M. (1980) Biochemistry and genetics of the ABO, Lewis, and P blood group systems (H. Harris and K. Hirschhorn, eds.), pp.1-136. Plenum Press, New York.
2. Hakomori, S. (1981) *Annu Rev Biochem* 50, 733-764.

3. Clausen, H., and Hakomori, S. (1989) *Vox Sang* 56, 1-20.
4. Clausen, H., White, T., Takio, K., Titani, K., Stroud, M., Holmes, E., Karkov, J., Thim, L., and Hakomori, S. (1990) *J Biol Chem* 265, 1139-1145.
5. Yamamoto, F., Clausen, H., White, T., Marken, J., and Hakomori, S. (1990) *Nature* 345, 229-233.
6. Clausen, H., Bennett, E.P., and Grunnet, N. (1994) *TCB* 2, 79-89.
7. Yamamoto, F., Marken, J., Tsuji, T., White, T., Clausen, H., and Hakomori, S. (1990) *J Biol Chem* 265, 1146-1151.
8. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4853-4857.
9. Navaratnam, N., Findlay, J.B., Keen, J.N., and Watkins, W.M. (1990) *Biochem. J.* 271, 93-98.
10. Yamamoto, F., McNeill, P.D., Yamamoto, M., Hakomori, S., Harris, T., Judd, W.J., and Davenport, R.D. (1993) *Vox Sang* 64, 116-119.
11. Joziassse, D.H., Shaper, N.L., Kim, D., Van den Eijnden, D.H., and Shaper, J.H. (1992) *J Biol Chem* 267, 5534-5541.
12. Larsen, R.D., Rivera-Marrero, C.A., Ernst, L.K., Cummings, R.D., and Lowe, J.B. (1990) *J Biol Chem* 265, 7055-7061.
13. Joziassse, D.H., Shaper, J.H., Jabs, E.W., and Shaper, N.L. (1991) *J Biol Chem* 266, 6991-6998.
14. Joziassse, D.H. (1992) Oxford University Press 271-277.
15. Henion, T.R., Macher, B.A., Anaraki, F., and Galili, U. (1994) *Glycobiol* 4, 193-201.
16. Ferguson-Smith, M.A., Aitken, D.A., Turleau, C., and Grouchy, J. (1976) *Hum Genet* 34, 35-43.
17. Kleene, R., and Berger, E.G. (1993) *Biochim. Biophys. Acta* 1154, 283-325.