Vol. 187, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 31, 1992 Pages 366-374

HUMAN HISTO-BLOOD GROUP A^2 TRANSFERASE CODED BY A^2 ALLELE, ONE OF THE A SUBTYPES, IS CHARACTERIZED BY A SINGLE BASE DELETION IN THE CODING SEQUENCE, WHICH RESULTS IN AN ADDITIONAL DOMAIN AT THE CARBOXYL TERMINAL

Fumi-ichiro Yamamoto*, Patricia D. McNeill and Sen-itiroh Hakomori

Biomembrane Institute, 201 Elliott Ave W, Seattle, WA 98119;
and Dept. of Pathobiology, University of Washington, Seattle, WA

Received July 17, 1992

SUMMARY: We have identified a possible mutation which characterizes A^2 alleles (a minor subtype of A) at the human histo-blood group ABO locus based on polymerase chain reaction (PCR) of genomic DNA, followed by nucleotide sequencing of the amplified fragments. The A^2 subtype has a single base deletion near the carboxyl terminal. As a result of frame-shifting, A^2 transferase possesses an extra domain. Introduction of this single base deletion into the A^1 transferase cDNA expression construct drastically decreased the A transferase activity in DNA-transfected HeLa cells.

Understanding of the histo-blood group ABO system has been a major focus in transfusion research. Since its discovery by Landsteiner in 1900 (1), the ABO system has been the best-studied human alloantigen system. Since mistakes in transfusion can have fatal results, reliable techniques to accurately identify blood types are crucial. Conventional analysis has been dependent on immunological detection of erythrocyte antigens using polyclonal and/or monoclonal antibodies against these antigens. The results of red cell grouping are usually confirmed by reverse (serum) grouping, i.e., testing the individual's serum with known A_1 and B erythrocytes.

We previously elucidated the molecular genetic basis of the ABO system by cloning of cDNA encoding A^1 transferase (2) based on the partial amino acid (a.a.) sequence of the purified enzyme (3), and subsequent cDNA cloning of B and O allelic cDNAs followed by nucleotide sequencing of the isolated cDNA clones (4). Availability of the nucleotide sequence has made genotyping at the ABO locus possible. Three major alleles (A^1, B, and O) have been successfully diagnosed by allele-specific restriction enzyme cleavage analysis and/or allele-specific oligodeoxynucleotide hybridization combined with PCR techniques (5).

*To whom correspondence should be addressed.

0006-291X/92 $4.00

We have now extended our study to the molecular basis of A^1 and A^2 subtypes. In 1911 two different A antigens were described by von Dungern (6). Serum from blood group B individuals contain two kinds of antibodies, anti-A and anti-A_1 which can be separated by absorption with erythrocytes from appropriate blood type individuals. The erythrocytes which react with only anti-A and not with anti-A_1 and those which react with both anti-A and anti-A_1 are classified as A_2 and A_1, respectively. Plant lectin from *Dolichos biflorus* has been used as an anti-A_1 reagent. The incidence of A_2 phenotype varies among different races; e.g., it is quite common in Caucasians (20% of the A population), but rare in Japanese (less than 1%).

The chemical nature of A_1/A_2 differences has been clarified during the past decade. The immunodominant structure of A antigen is GalNAcα1→3(Fucα1→2)Gal. In the A_1 phenotype, this structure is found on types I (Galβ1→3GlcNAcβ1→R), II (Galβ1→4GlcNAcβ1→R), III (Galβ1→3GalNAcα1→R), and IV (Galβ1→3GalNAcβ1→R) core structures. Types III and IV are also called repetitive-A and globo-A, respectively. However, this immunodominant structure in the A_2 phenotype is found only on types I and II core structures, and is less abundant (7,8).

In comparison of A^1 and A^2 transferases using lacto-N-fucopentaose I and 2'-fucosyllactose as acceptors, the enzymes were found to differ in their cation requirements, pH optima, and Km values; it was concluded that the A^1 and A^2 transferases differ qualitatively (9). A^1 transferase in serum had an isoelectric point (in the pH range 9.5-10) similar to that of the enzyme in cyst fluids; however, A^2 transferase in serum had a considerably lower isoelectric point in pH range 6-7. Different pI values of A^1 and A^2 transferases in the serum of a donor with genotype A^1A^2 allowed preparative separation of the two enzymes by isoelectric focusing (10). Post-translational modification of the enzyme protein has been suggested to explain this dissimilarity in pI values of the A^2 transferases in ovarian cyst fluids and serum samples (10).

Here we report the identification of a possible mutation in the nucleotide sequence of the A^2 gene which seems to be responsible for restricted substrate usage and weaker A transferase activity of A^2 transferase.

MATERIALS AND METHODS

Materials: Proteinase K was obtained from BRL (Gaithersburg, MD). Taq DNA polymerase for PCR was purchased from Perkin-Elmer Cetus (Norwalk, CT). Oligodeoxynucleotides were synthesized with Applied Biosystems 380B DNA synthesizer. Geneclean Kit was from BIO 101 (La Jolla, CA). Sequencing vector pT7T3U18 was purchased from Pharmacia-LKB (Piscataway, NJ). *E.coli XL-1-blue* frozen competent cells were from Stratagene (La Jolla, CA) and α-^{32}P dATP was from Amersham Corp.

(Arlington Heights, IL). Buffy coat fraction of human blood was obtained from American Red Cross, Portland, Oregon.

Purification of genomic DNA: Ten ml of buffy coat fraction was washed twice with the same volume of PBS and used for genomic DNA preparation by Proteinase K-SDS method (11).

Polymerase chain reaction and nucleotide sequencing: PCR was performed per manufacturer's protocol, using a DNA-Thermal Cycler from Perkin-Elmer Cetus. One μg of genomic DNA was used as a template for the amplification reaction. The pairs of synthetic oligodeoxynucleotide primers used for subcloning were: fy-121 and fy-122, fy-101 and fy-104, fy-111 and fy-112. The nucleotide sequences of these oligos are:

fy-101; CGGAATTCCCGTCCGCCTGCCTTGCAG,
fy-104; CGGAATTCGAAATCGCCCTCGTCCTT,
fy-111; CGGAATTCGCTGGAGGTGCGCGCCTAC,
fy-112; CGGAATTCCGTTCTGCTAAAACCAAG,
fy-121; CGGAATTCCATGTGACCGCACGCCT,
fy-122; CGGAATTCTCTACCCTCGGCCACC.

All the Eco RI sites are artificial. Five μl each of 20pmol/μl oligos was added to the heat-denatured DNA, followed by addition of the reaction mixture (45 μl H_2O, 10 μl 10x reaction buffer, 16 μl 1.25mM dNTP mixture, and 1 μl 2.5 units/μl Taq DNA polymerase). Two drops of paraffin oil were overlaid to prevent evaporation. Amplification was performed by step-cycle mode of 40 rounds of 94°C 2 min, 50°C 2 min, and 70°C 3 min, followed by one round of 94°C 2 min, 50°C 3 min, and 70°C 10 min, and samples were left at 10°C until processing occurred. Amplified DNA was extracted with 100 μl phenol:chloroform:isoamyl alcohol mixture (25:24:1), and the aqueous fraction was transferred into Eppendorf tubes with 12 μl 3M sodium acetate(pH 7.5) and 250 μl ethanol. After centrifugation, the pellet was dried and resuspended in H_2O and subjected to the restriction enzyme digestion with Eco RI. DNA was then electrophoresed through 2% agarose gel for size fractionation, recovered from gel fragments with Geneclean Kit, ligated with Eco RI-digested, BAP-treated pT7T3U18 sequencing vector, and used for DNA transformation of *E. coli XL-1-blue* strain competent cells. DNA from transformant clones was prepared in a mini-scale and analyzed for the inserts. DNA from a multiple number (more than 10) of correct constructs was prepared in large-scale, alkaline-denatured, and used for nucleotide sequencing by Sanger dideoxy termination method (12).

Introduction of a single base deletion into the A^1 transferase cDNA expression construct: A genomic DNA fragment from an A_2 individual was amplified by PCR as described above using two primers, fy-48; TACTACCTGGGGGGGTTCTT, and fy-78; CCGGATCCGTGTGATTTGAGGTGGGGAC (Bam HI site is artificial), purified, and subjected to restriction enzyme digestion with Ava I and Bam HI. After electrophoresis, the gel fragment containing Ava I/ Bam HI 251 bp DNA fragment was excised and DNA was extracted and ligated with Ava I/ Bam HI vector fragment from A^1 transferase cDNA expression construct p59-5/66-7 (s) (13).

DNA transfection and immunodetection of expression: DNA transfection was performed as previously described (13) using HeLa cells as the recipient by $CaPO_4$ method. Three days after transfection, cells were recovered and subjected to immunostaining and FACS analysis. The primary antibodies used were anti-A and anti-B murine monoclonal antibody mixtures (Ortho Diagnostics, Raritan, NJ). FITC-conjugated goat anti-mouse immunoglobulin was used as the secondary antibody.

RESULTS

In order to clarify the molecular basis for A^1 and A^2 subtypes, we employed a PCR approach. Because of the unavailability of cell lines with A_2 phenotype, genomic DNA from A_2 individuals was used as a template for PCR amplification. The protein-coding nucleotide sequence of A^2 subtype was determined for the last two coding exons by sequencing multiple numbers of independent transformant clones. Two pairs of primers (fy-101 and fy-104, and fy-111 and fy-112) were used to amplify the genomic DNA containing the last coding exon. This exon contains about 65% of the coding sequence of the membrane-bound form of A^1 and B transferases. The corresponding sequence for primer fy-101 resides in the intron preceding this exon (the last base of fy-101 is located just adjacent to the splicing acceptor site), and that of fy-112 resides in the 3' untranslated region (the last base of fy-112 is located 53 bases away from the 3' side of the termination codon). Primers fy-111 and fy-104 are homologous to the sequences in the last exon except for their artificial Eco RI sequences, and the amplified fragments by these oligo pairs overlap each other. The primers fy-121 and fy-122 were used to amplify the genomic DNA in such a way that the preceding exon would be covered (the last bases of fy-121 and fy-122 are located 16 bases away from the 5' side of the splicing acceptor site and 11 bases away from the 3' side of the splicing donor site, respectively). This exon contains 13% of the coding sequence of the membrane-bound form of these transferases. Most of the coding region of the soluble form of these transferases (91%; 274 a.a.s out of 301) is contained in these two exons.

Comparison of the deduced a.a. sequence and the partial nucleotide sequences containing the important difference, are shown in Figs. 1 and 2. Two differences have been identified; one single base substitution and one single nucleotide deletion. The nucleotide substitution (T in A^2 and C in A^1 at nucleotide position 467 counting from the A residue of the initiation codon) results in the a.a. substitution (leucine in A^2 transferase and proline in A^1 transferase at a.a. position 156). The same single base substitution was previously found in FY-59-5 cDNA clones (2,4), and the resulting a.a. substitution was found to be incapable of drastically altering enzymatic activity or sugar-nucleotide donor specificity, based on the observed expression of chimeric cDNAs in transfected HeLa cells (13). The single nucleotide deletion was found in three stretches of Cs (nucleotide position 1059-1061 in A^1 allele). This single base deletion, located at the end of the C-terminal, changes the reading frame and results in a protein with 21 additional a.a.s (Figs. 1, 2).

As the next step to determine whether this single base deletion is common among A^2 alleles, DNA from buffy coat fractions of seven more A_2 individuals was analyzed by

```
A1      MAEVLRTLAGKPKCHALRPMILFLIMLVLVLFGYGVLSPR   40
A2      ??????????????????????????????????????

A1      SLMPGSLERGFCMAVREPDHLQRVSLPRMVYPQPKVLTPC   80
A2      ??????????????????????????????????????

A1      RKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTV  120
A2      ****************************************

A1      FAIKKYVAFLKLFLETAEKHFMVGHRVHYYVFTDQPAAVP  160
A2      ********************************L*******

A1      RVTLGTGRQLSVLEVRAYKRWQDVSMRRMEMISDFCERRF  200
A2      ****************************************

A1      LSEVDYLVCVDVDMEFRDHVGVEILTPLFGTLHPGFYGSS  240
A2      ****************************************

A1      REAFTYERRPQSQAYIPKDEGDFYYLGGFFGGSVQEVQRL  280
A2      ****************************************

A1      TRACHQAMMVDQANGIEAVWHDESHLNKYLLRHKPTKVLS  320
A2      ****************************************

A1      PEYLWDQQLLGWPAVLRKLRFTAVPKNHQAVRNP=
A2      ********************************RERLPGA  360

A1
A2      LGGLPAAPSPSRPWF=                          375
```

Figure 1. Comparison of deduced amino acid sequences for A^1 vs. A^2 transferases

The initiation codon of membrane-bound forms of A^1 transferase is numbered 1 for deduced a.a. sequence. Soluble forms of enzymes start at alanine at a.a. 54. The a.a. at position 156 (marked in bold type) may not be A^2-specific. *, identical a.a. as shown above. =, termination codon. ?, unidentified a.a.

```
A1      AACCACCAGGCGGTCCGGAACCCGTGAGCGGCT    1071
A2      ********************-**********    1070
A1       N  H  Q  A  V  R  N  P  =            354
A2       *  *  *  *  *  *  *  R  E  R  L     357

A1      GCCAGGGGCTCTGGGAGGGCTGCCGGCAGCCCC    1104
A2      ********************************    1103
A1
A2        P  G  A  L  G  G  L  P  A  A  P     368

A1      GTCCCCCTCCCGCCCTTGGTTTTAG            1129
A2      ************************             1128
A1
A2        S  P  S  R  P  W  F  =               375
```

Figure 2. Comparison of nucleotide and deduced amino acid sequences around the area of difference for A^1 vs. A^2 alleles

The initiation codon and its A residue of membrane-bound forms of A^1 transferase are numbered 1 for deduced a.a. and nucleotide sequences respectively. The symbol (—) indicates the position of the single base deletion (one of the Cs from 1059-1061). *, identical a.a. as shown above. =, termination codon.

TABLE I. FACS analysis of immunostained DNA-transfected cells

DNA	Positive cell % with anti-A	Positive cell % with anti-B	adjusted ratio
$pA^2(1)$	0.7	16.1	1.0
$pA^2(2)$	0.2	2.8	1.6
$pA^1(1)$	21.7	10.7	46.6
$pA^1(2)$	22.8	10.2	51.4
no DNA	0.0	0.0	—

DNA preparations from two different clones (1 and 2) of each construct were used for DNA transfection experiments. Plasmid pA^2 has the single base deletion while plasmid pA^1 does not; otherwise they have identical nucleotide sequences. Taking into consideration that the efficiency of DNA transfection differs among samples, judged by the different positive cell % with anti-B antibody, the relative positive cell % with anti-A antibody was calculated and shown under "Adjusted ratio."

PCR and nucleotide sequencing. The single nucleotide deletion was found in all samples. However, the same nucleotide substitution at nucleotide 467 was also observed in all samples.

To confirm that the frameshift mutation indeed gives rise to an enzyme with kinetic characteristics corresponding to those of the A^2 transferase, this single nucleotide deletion was introduced into an A^1 transferase cDNA expression construct by a PCR approach. I.e., we replaced the Ava I/ Bam HI fragment of p59-5/66-7 (s) (13) with PCR-amplified Ava I/ Bam HI fragment from an A_2 individual whose genotype is A^2O. This replacement resulted in a decrease of 22 bp because of the size difference of the 3' untranslated region in these fragments. The nucleotide sequences of several recombinant clones were determined. Some of the clones had a single base deletion while others did not, which represented either A^2 or O allele. These two types of clones differed only in this single base deletion. Both possessed the single base substitution because the backbone construct was derived from FY-59-5. Therefore, these two constructs were compared in terms of their expression in the DNA-transfected HeLa cells. In order to standardize the efficiency of DNA transfection, 5 μg of pBBBB plasmid DNA (13) was added to 15 μg of test DNA and used for DNA transfection. In addition, DNA

preparations from two different clones from each construct were tested (Table I). The introduction of the single base deletion decreased A transferase activity 30-50 fold.

DISCUSSION

There has been a long-standing debate on the quantitative and qualitative differences between A_1 and A_2 antigens. Initially, only quantitative differences were reported (14). The presence of an A_1-specific antibody was later thought to suggest the presence of A_1-specific structures (15). It was suggested that A antigens reside on both type I and type II core structures in A_1, but only type II in A_2 (16). Later, A_1 erythrocytes were found to express immunodominant A structure on type I, II, III, and VI core structures, whereas A_2 erythrocytes have it only on type I and II, and the number of A structures is much less (7,8,17). The qualitative difference between A_1 and A_2 antigens was explained by the weaker activity and restricted acceptor substrate specificity of A^2 transferase.

When we elucidated the molecular basis of the allelic nature for three major alleles, A^1, B, and O (4), we anticipated that a similar mechanism (i.e., a.a. substitutions to distinguish A and B alleles) might be the basis for the differences between A^1 and A^2 transferases. Unexpectedly, a common single base deletion was found to characterize the A^2 allele. Because all the A_2 individuals tested were found to be heterozygous (A^2O) based on the presence of an O-allele-specific single base deletion at nucleotide position 261 in the O allele, and the genomic fragments containing these two deletions were amplified in two separate fragments, we cannot exclude the possibility that in some cases one allele (O) has both deletions, and the A^2 allele has nucleotide difference(s) somewhere else. However, the simpler explanation is that the single base deletion at the C-terminal is a common feature for A^2 alleles. A single base deletion was previously found to characterize the non-functional O allele (4). While the single base deletion in O alleles is located close to the N-terminal, that of the A^2 allele is close to the C-terminal. As a result of frame-shift, the O allele is unable to code for the functional glycosyltransferase, but the A^2 allele is able to code for the protein with 21 additional a.a.s. While this glycosyltransferase is still functional, there are changes in activity strength and substrate specificity.

We have previously shown that a single a.a. substitution may change sugar-nucleotide donor specificity, based on chimera constructions and DNA transfection experiments (13). The location of the a.a. substitutions suggests the importance of this region for interaction with sugar-nucleotide molecules. The present findings suggest that the C-terminal may be important for the strength and acceptor specificity of enzymatic

activity, although this could alternatively be due to mainly the steric hindrance of the additional 21 a.a.s.

The explanation for the observation by Watkins et al. (10) that A^2 transferase in serum has acidic pI remains obscure. Post-translational modification was suggested by these authors because A^2 transferase in cyst fluids shows the same pI as A^1 transferase. Although we could not identify the possible N-glycosylation site (Asn-xxx-Ser/Thr) in the deduced a.a. sequence of the additional domain, there are two serine residues in that region, some of which may be O-glycosylated when secreted into serum.

Because we have not determined the nucleotide sequence of the entire coding region of the A^2 transferase, a possibility still remains that more differences in the coding sequence may exist. In addition, because we have not determined the nucleotide sequence of the cDNA, differences other than the coding region, for example, differences affecting the splicing pattern, may also exist. However, we do not think that this is the case because of the following reasons; 1). All the glycosyltransferases so far cloned are suggested to have a characteristic topology in the Golgi apparatus consisting of a short NH_2-terminal cytoplasmic tail, a signal-anchor domain which spans the membrane, an extended stem region, and a large COOH-terminal catalytic domain oriented within the lumen of the Golgi cisternae (18). About 78% of the entire coding region of membrane-bound form and 91% of soluble form A^1 and B transferases reside in the last two coding exons. Therefore this model fits well with the assumption that the catalytic domains of A and B are made up of by an a.a. sequence coded in the last coding exon and/or the preceding exon. 2). Another glycosyltransferase, α1→3galactosyltransferase, transfers galactose to the acceptor substrates by an α1→3 linkage as B transferase but utilizes the substrates without fucose linked to galactose by α1→2 linkage. Nucleotide and deduced amino acid sequences in the last two coding exons of A^1 and B transferases show the extensive sequence homology with murine (19) and bovine (20) α1→3galactosyltransferase genes but the sequence preceding these exons shows no homology at all (13). Similarity in enzymatic specificity and sequence homology suggest the occurrence of these genes from the same ancestral gene by gene duplication and divergence (21). Sequence information necessary for enzymatic function should have been conserved during evolution. Taken together with our experimental results from DNA transfection, we believe that the single base deletion found in the last coding exon is the crucial difference which discriminates A^2 allele from A^1 allele resulting in two transferases with different characteristics.

Our success in elucidating the molecular basis of the A^2 allele by PCR and nucleotide sequencing of the amplified fragments suggests that the same approach may

be useful when applied to other subtypes and rare polymorphisms of the human histo-blood group ABO system.

ACKNOWLEDGMENTS

We thank Miyako Yamamoto for technical assistance, Dr. Manjula Rystrom for assistance in FACS analysis, and Jennifer Stoeck and Dr. Stephen Anderson for scientific editing and preparation of the manuscript. This study was supported by funds from The Biomembrane Institute, in part under a research contract from Otsuka Pharmaceutical Co.

REFERENCES

1. Landsteiner, K. (1900) Zbl. Bakt. **27**, 357-363.
2. Yamamoto, F., Marken, J., Tsuji, T., White, T., Clausen, H., and Hakomori, S. (1990) J. Biol. Chem. **265**, 1146-1151.
3. Clausen, H., White, T., Takio, K., Titani, K., Stroud, M.R., Holmes, E.H., Karkov, J., Thim, L., and Hakomori, S. (1990) J. Biol. Chem. **265**, 1139-1145.
4. Yamamoto, F., Clausen, H., White, T., Marken, J., and Hakomori, S. (1990) Nature **345**, 229-233.
5. Saiki, R.K., Scharf, S.J., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985) Science **230**, 1350-1354.
6. Dungern,E.v. and Hirszfeld,L. (1911) Z. Immun. Forsch. **8**, 526-562.
7. Clausen, H., Watanabe, K., Kannagi, R., Levery, S.B., Nudelman, E., Arao-Tomono, Y., and Hakomori, S. (1984) Biochem. Biophys. Res. Commun. **124**, 523-529.
8. Clausen, H., Levery, S.B., Nudelman, E., Tsuchiya, S., and Hakomori, S. (1985) Proc. Natl. Acad. Sci. USA **82**, 1199-1203.
9. Schachter, H., Michaels, M.A., Tilley, C.A., Crookston, M.C., and Crookston, J.H. (1973) Proc. Natl. Acad. Sci. USA **70**, 220-224.
10. Topping, M.D. and Watkins, W.M. (1975) Biochem. Biophys. Res. Commun. **64**, 89-96.
11. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, NY.
12. Sanger, F., Miklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
13. Yamamoto, F. and Hakomori, S. (1990) J. Biol. Chem. **265**, 19257-19262.
14. Makela, O., Ruoslahti, E., and Ehnholm, C. (1969) J. Immunol. **102**, 763-771.
15. Mohn, J.F., Cunningham, R.K., and Bates, J.F. (1977) in Human blood groups (Mohn, J., Plunkett, R., Cunningham, R., and Lambert, R., eds.), pp. 316-325. Karger, New York.
16. Kisailus, E.C. and Kabat, E.A. (1978) J. Exp. Med. **147**, 830-843.
17. Breimer, M.E. and Jovall, P.-A. (1985) FEBS Lett. **176**, 165-172.
18. Paulson, J.C. and Colley, K.J. (1989) J. Biol. Chem. **264**, 17615-17618.
19. Joziasse, D.H., Shaper, J.H., van den Eijnden, D.H., Van Tunen, A.J., and Shaper, N.L. (1989) J. Biol. Chem. **264**, 14290-14297.
20. Larsen, R.D., Rajan, V.P., Ruff, M.M., Kukowska-Latallo, J., Cummings, R.D., and Lowe, J.B. (1989) Proc. Natl. Acad. Sci. USA **86**, 8227-8231.
21. Yamamoto, F., McNeill, P.D., and Hakomori, S. (1991) Biochem. Biophys. Res. Commu. **175**, 986-994.