Ann. Hum. Genet. (2002), **66**, 1–27 © *University College London*
DOI: 10.1017/S0003480001008995 *Printed in the United Kingdom*

Sequence variation at the human *ABO* locus

S. P. YIP
Department of Nursing and Health Sciences, The Hong Kong Polytechnic University, China

(Received 11.7.01. Accepted 10.10.01)

SUMMARY

The ABO blood group is the most important blood group system in transfusion medicine. Since the *ABO* gene was cloned and the molecular basis of the three major alleles delineated about 10 years ago, the gene has increasingly been examined by a variety of DNA-based genotyping methods and analysed in detail by DNA sequencing. A few coherent observations emerge from these studies. First, there is extensive sequence heterogeneity underlying the major *ABO* alleles that produce normal blood groups A, B, AB and O when in correct combination with other alleles. Second, there is also extensive heterogeneity underlying the molecular basis of various alleles producing ABO subgroups such as A_2, A_x and B_3. There are over 70 *ABO* alleles reported to date and these alleles highlight the extensive sequence variation in the coding region of the gene. A unifying system of nomenclature is proposed to name these alleles. Third, extensive sequence variation is also found in the non-coding region of the gene, including variation in minisatellite repeats in the 5′ untranslated region (UTR), 21 single nucleotide polymorphisms (SNPs) in intron 6 and one SNP in the 3′ UTR. The haplotypes of these variations reveal a specific relationship with the major *ABO* alleles. Fourth, excluding the common alleles, about half of the remaining alleles are due to new mutations and the other half can better be explained by intragenic recombination (both crossover and gene conversion) between common alleles. In particular, the recombination sites in hybrid alleles can be quite precisely defined through haplotype analysis of the SNPs in intron 6. This indicates that recombination is equally as important as point mutations in generating the genetic diversity of the *ABO* locus. Finally, a large number of *ABO* genotyping methods are available and are based on restriction analysis, allele specific amplification, mutation screening techniques or their combinations.

INTRODUCTION

The ABO blood group system (MIM number: 110300) was discovered by K Landsteiner a century ago. It is the most important blood group system in transfusion medicine. The serological and genetic characteristics of the system and the biosynthesis of its antigens have been well established (Daniels, 1995; Watkins, 1980).

Correspondence: S. P. Yip, Department of Nursing and Health Sciences, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China. Fax: (852) 2364 9663.
E-mail: hsspyip@polyu.edu.hk

The antigenic determinants of the system are oligosaccharides found on glycoproteins and glycolipids. As such they are not direct gene products of the *ABO* gene. Instead, the *ABO* gene encodes enzymes known as glycosyltransferases which transfer specific sugar residues to a precursor substance (the H antigen) to produce the A and B antigens. There are three major alleles at the *ABO* locus: alleles *A*, *B* and *O*. The *A* allele encodes $\alpha 1 \rightarrow 3$ N-acetylgalactosaminyl-transferase (A transferase; EC 2.4.1.40) which adds N-acetylgalactosamine (GalNAc) to the H antigen to form the A antigen. The *B* allele encodes $\alpha 1 \rightarrow 3$ galactosyltransferase (B transferase; EC 2.4.1.37) which transfers galactose

(Gal) to the H antigen to construct the B antigen. Both transferases use the same acceptor substrate (H antigen) but different nucleotide-sugar donor substrates (uridine diphosphate-GalNAc or UDP-GalNAc for A transferase and UDP-Gal for B transferase). On the other hand, the *O* allele does not produce an active enzyme.

THE HUMAN *ABO* GENE

Cloning of the ABO gene

The complementary deoxyribonucleic acid (cDNA) encoding the human A transferase was cloned and sequenced by Yamamoto *et al.* (1990*b*). On the basis of the partial amino acid sequence determined for a purified A transferase preparation (Clausen *et al.* 1990), degenerate primers were prepared to amplify a DNA fragment using polymerase chain reaction (PCR). The DNA fragment was used as a probe to screen a cDNA library constructed using poly(A)$^+$ ribonucleic acid (RNA) from a human stomach cancer cell line known to express high levels of A antigen. Nucleotide sequence analysis of the isolated positive clones revealed a coding region of 1065 basepairs (bp) encoding a protein of 41 kiloDalton with 354 amino acids (Yamamoto *et al.* 1990*b*). Hydrophobicity plot analysis reveals three domains in the encoded transferase: an N-terminal short stretch, a transmembrane hydrophobic region and a long C-terminal domain. The long C-terminal domain most likely contains the catalytic domain because the catalytically active, purified soluble form of the enzyme lacks the N-terminal and the hydrophobic domains. This conforms to the common general structure of all glycosyltransferases, which are type 2 transmembrane glycoproteins having a large catalytic domain facing the lumen of the Golgi apparatus (Colley, 1997).

The genomic organization of the human *ABO* locus was determined independently and more or less simultaneously by two separate groups. One group used a PCR probe to screen a human P1 library and identified a positive phage clone (Bennett *et al.* 1995). Another group used A transferase cDNA probes to screen a human leucocyte genomic library and then a human placenta genomic library, and obtained multiple overlapping clones containing the *ABO* gene (Yamamoto *et al.* 1995). The human *ABO* locus spans over 18 kilobases (kb) and consists of seven exons. All exon/intron boundaries conform to the GT-AG consensus rule. The exons range in size from 28 to 688 bp. The transmembrane region (amino acid residues 17–37) is mostly present in exon 2, and the N-terminal of the soluble protein (amino acid residue 54) is found in exon 4. The two largest exons (6 and 7) encode 77 % of the full coding region and 91 % of the catalytically active soluble transferase protein. Fluorescent *in situ* hybridization to metaphase chromosomes using the isolated P1 phage localized the *ABO* locus to 9q34 (Bennett *et al.* 1995), which is in agreement with the previously reported linkage analysis studies (Ferguson-Smith *et al.* 1976).

Regulation of gene expression

The promoter region of the *ABO* gene contains a CpG island whose methylation status correlates well with the level of gene expression in tested cell lines (Kominato *et al.* 1999). Sequence analysis of the 5′ upstream region of the *ABO* gene reveals the presence of several GC boxes (consensus sequence: GGGCGG in either orientation) just upstream of two possible transcription start sites, but neither TATA nor CAAT boxes are found close to these sites (Yamamoto *et al.* 1995). Specifically, the region between −117 and +31 has promoter activity and is sufficient to direct the expression of a reporter when placed 5′ adjacent to the reporter gene (Kominato *et al.* 1997). A CpG island extends from the immediate 5′ flanking region through the first exon and into the first intron (from −0.7 kb to +0.6 kb): this region has a G+C content of 76 %, a CpG density of 11.7 % and a CpG/GpC ratio of 0.9 (Kominato *et al.* 1999). The CpG island was found to be hypomethylated in cell lines expressing the *ABO* gene, but hypermethylated in cell lines not

expressing the gene. Although transient transfection of reporter constructs containing *ABO* gene promoter sequence demonstrated constitutive transcriptional activity of the *ABO* gene promoter, the promoter activity was found to be suppressed if the promoter region was methylated *in vitro* using *Hha*I methylase before transfection into the expressor gastric cancer cell line KATO III cells. On the other hand, demethylation of the *ABO* gene promoter by treating the non-expressor gastric cancer cell line MKN28 cells with DNA methylase inhibitor resulted in appearance of A transferase transcripts and A antigens synthesised by A transferase. These results suggest that expression of the *ABO* gene is dependent on the DNA methylation status of the promoter region.

Transcription of the *ABO* gene depends on the binding of transcription factor CBF/NF-Y to an upstream minisatellite sequence (Kominato *et al.* 1997). A minisatellite is found at positions −3843 to −3672 relative to the upstream transcription start site and composed of four tandem copies of a 43 bp repeat unit. This minisatellite was found to have an enhancing effect on transcription when introduced into a promoterless luciferase reporter vector and analysed by transient transfection assays with gastric cancer cell line KATO III cells. The transcription factor CBF/NF-Y was shown to bind to the 43 bp repeat unit in the minisatellite by electrophoretic mobility shift assay. CBF/NF-Y binds specifically to the CCAAT sequence, a common DNA motif found in the promoter of numerous mammalian genes transcribed by RNA polymerase II (reviewed in Maity & de Crombrugghe, 1998). The CCAAT motif is found towards the 5′ end of the 43 bp repeat. Artificially introduced mutation in the CCAAT motif of the minisatellite was found to abrogate the binding of CBF/NF-Y. Functional importance of these binding sites in enhancer activity of the minisatellite was further confirmed by transfection experiments using reporter constructs with mutated binding sites. Thus, transcription regulation of the human *ABO* gene is dependent on the binding of CBF/NF-Y to the minisatellite sequence.

THREE MAJOR *ABO* ALLELES RESULTING FROM SEQUENCE VARIATION IN THE CODING SEQUENCE

There are three major *ABO* alleles: *A*, *B* and *O* (also known as *A101* or A^1, *B101*, and *O101*, O^1 or *O01*, respectively; Figs 1–3). Single base deletion and single base substitutions account for the differences among these three major alleles. These base changes occur in the two largest exons (6 and 7) of the *ABO* gene. Their molecular basis was delineated by cloning and sequencing of allelic cDNAs. The *A101* allele is usually used as the reference against which all other alleles are compared although the cDNA sequence of the *A102* allele was reported first (Yamamoto *et al.* 1990*b*).

The *O101* allele differs from the *A101* allele by a single base (G) deletion at nucleotide (nt) 261 corresponding to amino acid 87 of the A transferase (Fig. 3; Yamamoto *et al.* 1990*a*). This shifts the reading frame of the coding sequence and generates a premature termination codon downstream from the deletion. It is predicted to produce an altered and shortened polypeptide of 116 amino acids that lacks the C-terminal catalytic domain and hence is enzymatically inactive.

The *B101* allele, on the other hand, differs from the *A101* allele by seven single base substitutions within the coding sequence at nt 297, 526, 657, 703, 796, 803 and 930 (Fig. 2; Yamamoto *et al.* 1990*a*). Four of these base substitutions (nt 526, 703, 796 and 803) result in amino acid substitutions (residues 176, 235, 266 and 268). The respective amino acid residues at these four positions are arginine, glycine, leucine and glycine in A transferase, and glycine, serine, methionine and alanine in B transferase. In other words, these four amino acid substitutions explain all the differences in the activity and the nucleotide-sugar donor specificity of the A and B transferases.

The functional role of these four amino acids was studied by transient transfection and expression in HeLa cells of a series of artificial chimeric cDNA constructs in a plasmid expression vector (Yamamoto & Hakomori, 1990).

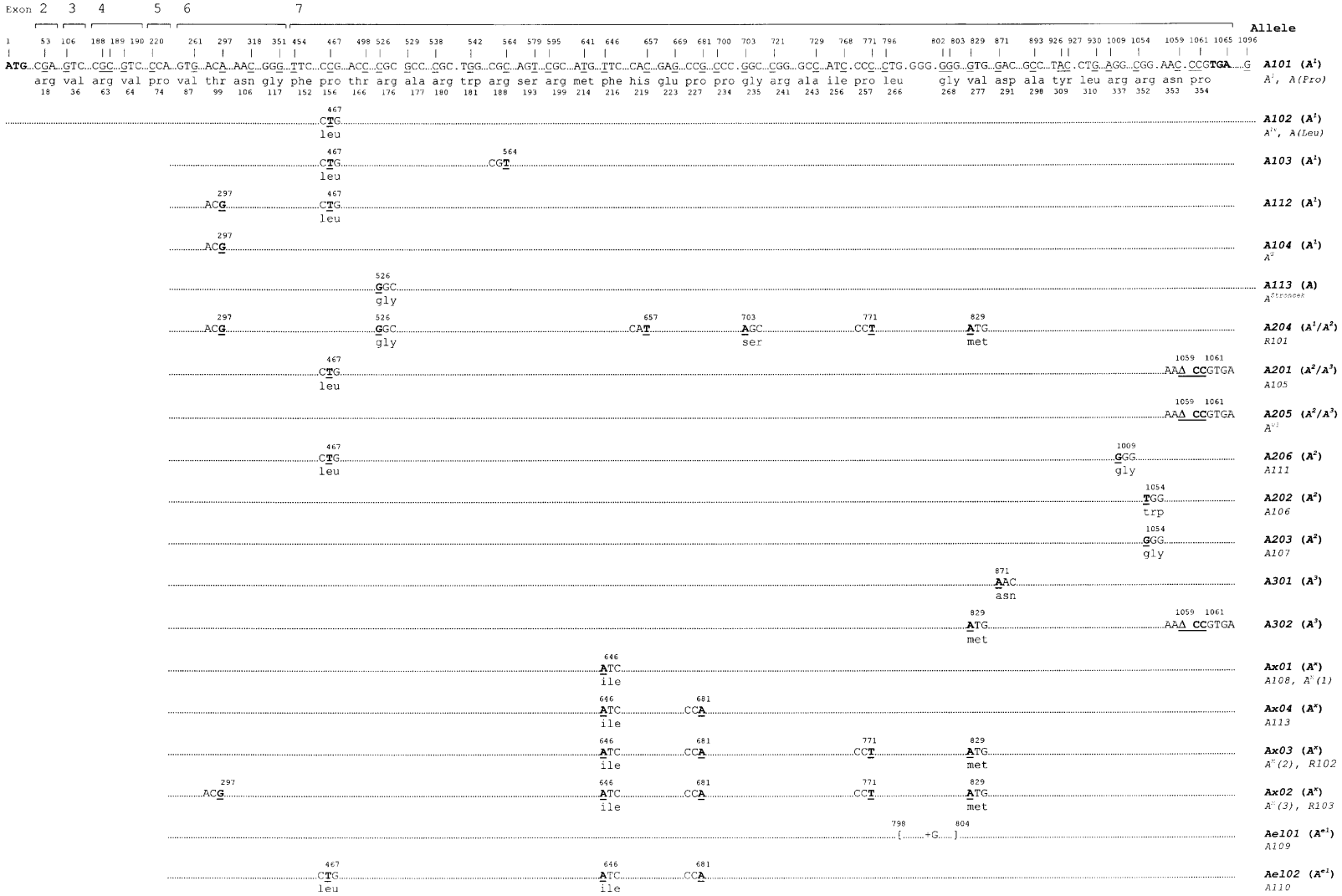

Fig. 1. Schematic comparison of nucleotide and deduced amino acid sequences of *A* alleles reported to date. Twenty *A* alleles are compared. The *A101* allele is used as the reference for comparison and only differences from this sequence are indicated. The extent of a dotted line represents the region of the *ABO* gene that is examined for the given allele. The proposed nomenclature is printed in boldface, below which are the alternative names, if any, used by the discoverer of the allele or other authors. The specificity of the allele is indicated in brackets.

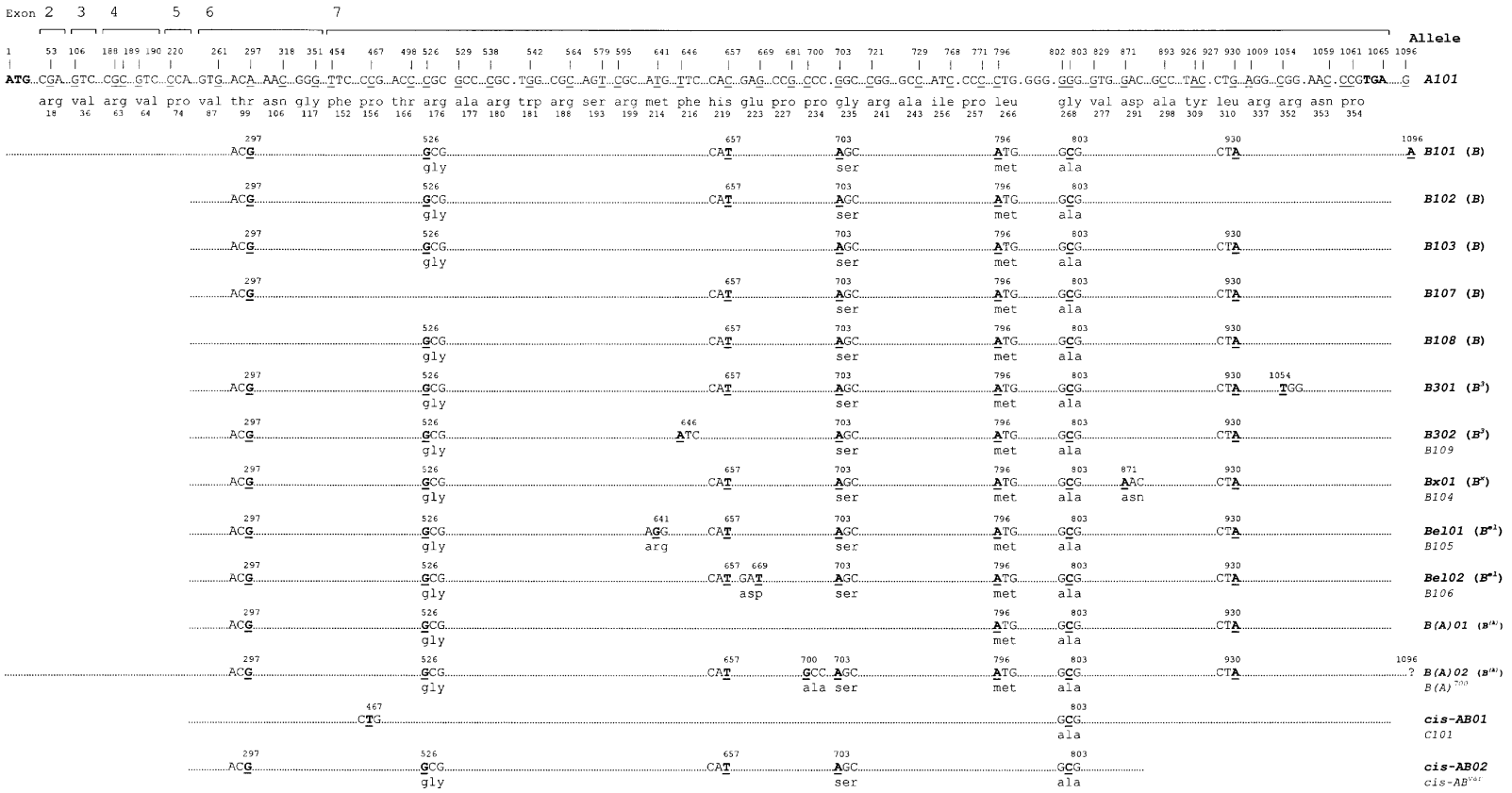

Fig. 2. Schematic comparison of nucleotide and deduced amino acid sequences of *B* alleles reported to date. Twelve *B* alleles and two *cis-AB* alleles are compared. For details, please refer to the legend of Figure 1.

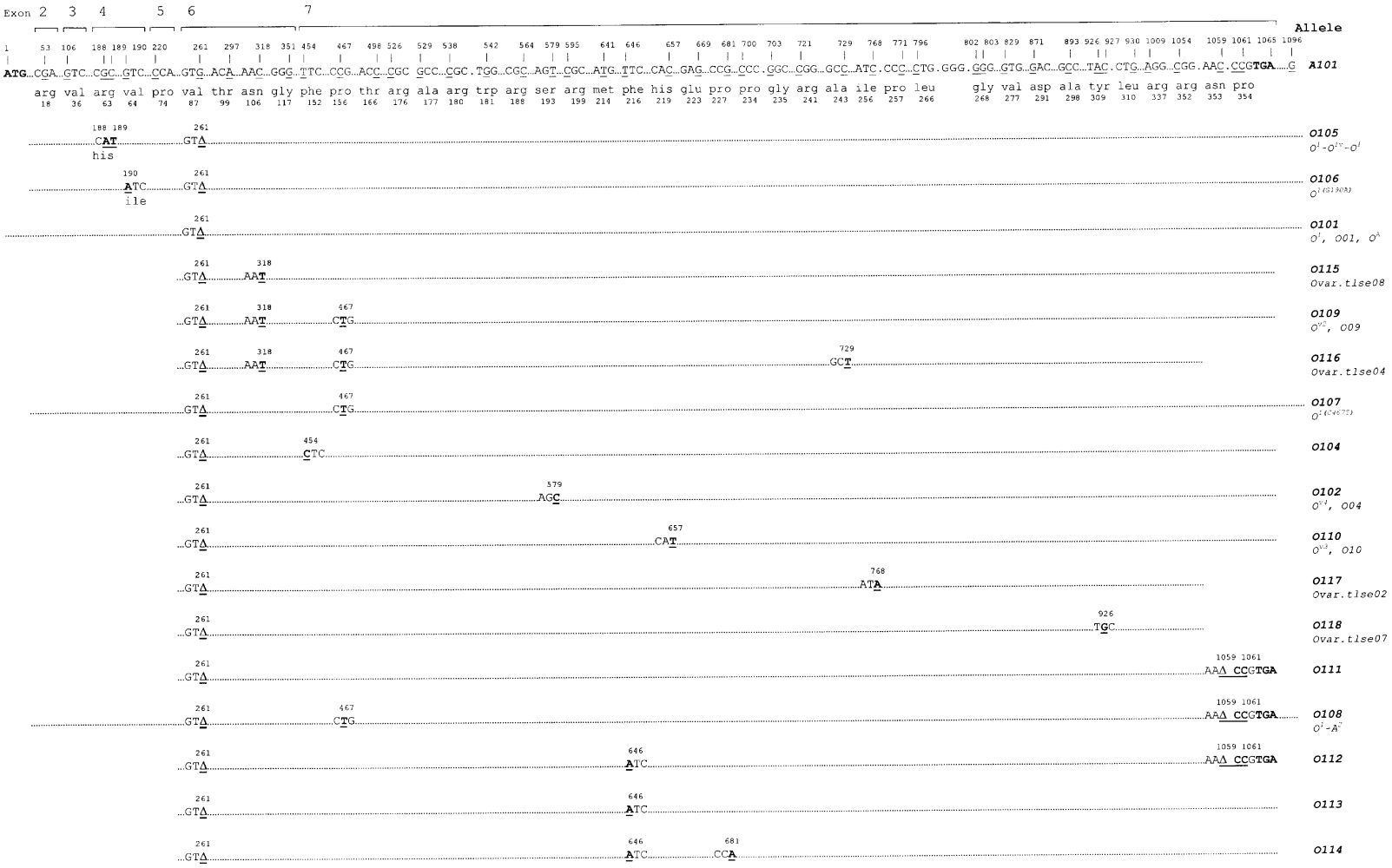

Fig. 3. For legend see facing page.

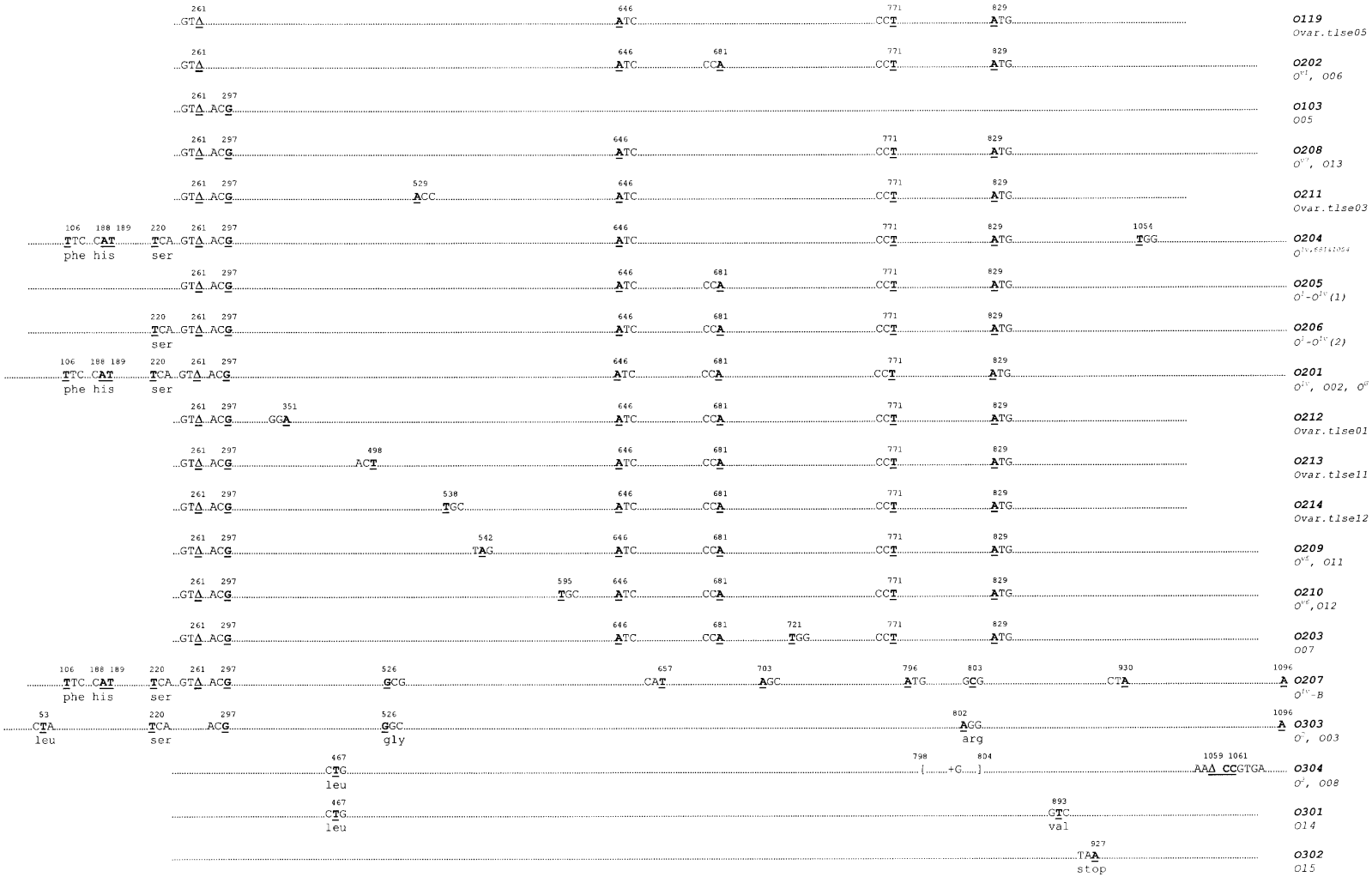

Fig. 3. Schematic comparison of nucleotide and deduced amino acid sequences of *O* alleles reported to date. Thirty-seven *O* alleles are compared. For details, please refer to the legend of Figure 1.

The transfected cells were analysed by immunostaining for the presence of cell surface A or B antigen. Untransfected HeLa cells were found to express the H antigen on the cell surface and shown to be of genotype *O/O* at the *ABO* locus. Sixteen chimeric cDNAs were constructed so that the encoded enzymes possessed, at each of the four amino acid positions, the residue present in either the A or B transferase. Transfection experiments with these chimeric constructs established that only A or B transferase activity was demonstrated when the amino acid residues at the third and the fourth positions (266 and 268) were leucine and glycine (i.e. AA) or methionine and alanine (i.e. BB), respectively. Both A and B transferase activities were detected when these two amino acid residues were methionine and glycine (i.e. BA), in this order. The results were dependent on the identity of the amino acid residue at the second position (235) when these last two positions were leucine and then alanine (i.e. AB). If the second position was glycine (A), only A transferase was observed. If it was serine (B), weak B transferase activity was detected in addition to the A transferase activity. Thus, these results show that the third and the fourth amino acid substitutions (leucine and glycine in A and methionine and alanine in B) are crucial in determining which nucleotide-sugar donor is utilized by the transferase. The second amino acid substitution (glycine in A and serine in B) also affects the nucleotide-sugar specificity. But the first amino acid substitution does not affect the specificity at all. These results were largely confirmed by a separate study that characterized the kinetics of wildtype A and B and hybrid A/B recombinant transferases (Seto *et al.* 1997). It is interesting to note that wildtype A and B transferases can utilize both nucleotide-sugar substrates but that the non-specific nucleotide-sugar substrate is utilized at an appreciably lower rate than the specific substrate.

The mechanistic role of the fourth amino acid substitution (residue 268, glycine in A and alanine in B) in donor substrate specificity was further explored using *in vitro* mutagenesis study. Substitution constructs were produced with every possible amino acid residue at this position and studied by immunodetection and *in vitro* enzymatic assay after transient transfection and expression in HeLa cells and COS 1 cells (Yamamoto *et al.* 1996). The results show that the size of the side chain of the amino acid at position 268, and partly its charge, are responsible for determining both the activity and the nucleotide-sugar donor specificity. This strongly suggests that this amino acid is directly involved in the recognition of, and binding to, the sugar moiety of the nucleotide-sugar substrates.

Extensive heterogeneity in sequence variation in the coding region

Since the elucidation of the molecular basis underlying the three major *ABO* alleles (Yamamoto *et al.* 1990*a*), the *ABO* locus has increasingly been examined by a variety of DNA-based genotyping methods (see below) and analysed in detail by direct DNA sequencing. Three coherent observations emerge from these studies. First, there is extensive sequence heterogeneity underlying the three major *ABO* alleles that produce the normal blood groups A, B, AB or O when in correct combination with other alleles. Second, there is also extensive heterogeneity underlying the molecular basis of various alleles producing ABO subgroups such as A_2, A_x and B_3. To date there are over 70 *ABO* alleles reported in the literature. Only exons 6 and 7 were examined for most of these alleles while the full coding sequence has been elucidated for seven alleles only, namely, *A101*, *A102*, *B101*, *B(A)02*, *O101*, *O201* and *O303*. Third, an increasing number of *ABO* alleles, usually individually rare, are found to have arisen or can better be explained by intragenic recombination at the *ABO* locus (i.e. recombination between two *ABO* alleles within the gene).

It should be mentioned at this point that there is up to now no consensus terminology for all the identified alleles, although there have been attempts to do so (Ogasawara *et al.* 1996*a*; Yamamoto, 2000). One of the reasons for this is

that new alleles are continuously being reported in the literature and the authors usually name new alleles in their own way. In this article, a standardised nomenclature is proposed which adopts the best features of the existing systems. For functional alleles *A* and *B*, the new system basically adopts the rules proposed by Yamamoto (2000) and examples can be found at http://www.bioc.aecom.yu.edu/bgmut/abo.htm. The name starts with the blood group or phenotype responsible, e.g. *A1* for group A_1, *A2* for group A_2, *A3* for group A_3, *B1* for group B, etc. This is followed by two digits usually indicating the order of discovery for a particular allele (e.g. *01*, *02*, etc.) within a group. The existing names will be used as far as possible unless there is discrepancy between a given name and the original report. New alleles will be named accordingly if they have not been included in the database mentioned above. This system has the advantage that the phenotype resulting from a given allele is obvious from the name. Note that there are rare occasions in which a given allele is found to produce more than one phenotype. For the non-functional alleles *O*, the new system adopts the major features proposed by Ogasawara *et al.* (1996*a*, 2001) with some modifications. *O* alleles are divided into three categories: *O101*-like, *O201*-like and non-deletional alleles (see below) with names starting with *O1*, *O2* and *O3*, respectively, and followed by two digits indicating the order of discovery within a category. The newly proposed names will be used in this article and alternative names used by the discoverers of the alleles or other authors will also be included wherever appropriate.

Heterogeneity of A alleles

A¹ alleles

Five *A* alleles have been reported to date that produce normal blood group A_1 with an *O* allele or with any one of themselves, and group A_1B with a normal *B* allele: *A101* (the classical A^1 allele), *A102* (or A^{IV}), *A103*, *A104* and *A112* (Fig. 1). These are the so-called A^1 alleles. When compared with the *A101* allele, allele *A102* has a nonsynonymous substitution at nt 467 of the cDNA sequence (467C>T) that changes the amino acid residue at codon 156 from proline to leucine (Pro156Leu) (Yamamoto *et al.* 1990*a*). This amino acid substitution does not seem to change the substrate specificity or the activity of the encoded A transferase, as reflected by the serum transferase activity (Ogasawara *et al.* 1996*b*) or the observed expression of cDNA constructs in transfected HeLa cells (Yamamoto & Hakomori, 1990). On the other hand, *A103*, *A104* and *A112* each have an additional synonymous base substitution on a different haplotype background. *A103* possesses 564C>T on the *A102* background while *A104* has 297A>G on the *A101* background and *A112* has 297A>G on the *A102* background (Ogasawara *et al.* 1996*a*, 2001). It is interesting to note that both *A104* and *A112*, with the synonymous substitution 297A>G, are identical to the *B* allele in exon 6 (see below). All three alleles are rare and have so far been reported in Japanese only. Both *A101* and *A102* are common, though in different populations. In Caucasian populations, *A101* is more common than *A102* (14–21% *vs.* 2–8%) (Nishimukai *et al.* 1996; Yip, 2000). However, in Oriental populations, *A101* is less common than *A102* (1–7% *vs.* 18–23%) (Fukumori *et al.* 1996*a*; Kang *et al.* 1997; Ogasawara *et al.* 1996*a*, 2001; Yip, 2000).

A113 (or $A^{Stroncek}$) has a nonsynonymous substitution 526C>G (Arg176Gly) in comparison with the *A101* sequence (Stroncek *et al.* 1995). The amino acid substitution is the first of four substitutions that mark the differences between A and B transferase. The allele predictably produces blood group A since this substitution does not affect the substrate specificity of the encoded transferase (Yamamoto & Hakomori, 1990).

A² alleles

A^2 alleles are responsible for blood group (i.e. phenotype) A_2 and are mainly found in Europe, the Near East and Africa (Daniels, 1995). This overall allele frequency usually does not exceed

10% in Caucasian populations and is very low (< 1%) in Oriental populations. There are six A^2 alleles (Fig. 1). *A201* (or *A105*) is the predominant A^2 allele in Caucasians and is characterised by the nonsynonymous substitution 467C>T (Pro156Leu) and a single C deletion in a series of three C's at nt 1059–1061 (hereafter designated as 1060delC) (Yamamoto *et al.* 1992). The resulting enzyme has 30–50 fold less A transferase activity, as shown by immunostaining of HeLa cells transfected with *A201*-construct. The single base deletion (1060delC) shifts the reading frame, produces a transferase with 21 additional amino acid residues at the C-terminal, and presumably is responsible for the reduction of the enzymatic activity. *A205* carries the 1060delC without the substitution 467C>T and was found in two Caucasians of group A_2 (Olsson & Chester, 1996*b*; Yip, 2000). With the presence of 1060delC as the only defining criterion, the A^2 allele frequencies range from 1.7% to 6.8% in various Caucasian populations (Gassner *et al.* 1996; Pearson & Hessner, 1998; Watanabe *et al.* 1997; Yip, 2000).

Three other A^2 alleles each have, near the 3′ end of the gene, a nonsynonymous substitution that results in amino acid substitution near the C-terminal of the encoded transferase: *A206* (or *A111*) with 1009A>G (Arg337Gly), *A202* (or *A106*) with 1054C>T (Arg352Trp) and *A203* (or *A107*) with 1054C>G (Arg352Gly) (Ogasawara *et al.* 1996*b*, 1998). Note that these base substitutions are new base changes not found in any of the common *ABO* alleles. On the other hand, *A204* (or *R101*) is a hybrid allele with four base substitutions (297A>G, 526C>G, 657C>T and 703G>A) characteristic of the *B* allele and two other substitutions (771C>T and 829G>A) characteristic of the *O201* allele (Ogasawara *et al.* 1996*b*). Three of these six base substitutions (nt 526, 703 and 829) are nonsynonymous and result in amino acid substitutions (Arg176Gly, Gly235Ser and Val277Met). Although the *A204*-encoded transferase has amino acid residues at 176 and 235 identical to those of the B transferase, those at positions 266 and 268 are identical to those of the A_1 transferase. On the basis of the results from transfection experiments with chimeric constructs (Yamamoto & Hakomori, 1990), this transferase is predicted to have A transferase activity. It seems likely that the amino acid substitution Val277Met reduces the enzymatic activity without affecting the nucleotide-sugar substrate specificity. Instead of the *A201* allele, these four alleles are the predominant A^2 alleles in Japanese (Ogasawara *et al.* 1996*b*, 1998). Furthermore, *A202* and *A203* are predominant in group A_2 individuals with the generic genotype A^2/O, while *A204* and *A206* are predominant in group A_2B individuals. Even more interesting is the finding that *A204* is expressed as group A_1 when in combination with an *O* allele, and gives group A_2B when combined with a *B* allele (Ogasawara *et al.* 1998). This is an example of the well-documented phenomenon of allelic competition (Daniels, 1995).

There are other interesting reports on the variation of phenotypes produced by A^2 alleles. An apparent *A201* allele was expressed as group A_x (or A_{weak}) when associated with an *O* allele in two Swedish families (Olsson & Chester, 1998). Interestingly, this *A201* allele behaved like a typical A^2 allele when associated with a *B* allele in one of these families. In another family, the allele *A205* was found to be expressed as group A_2 in a father genotyped as *A205/O114*, but as group A_3 in the two sons genotyped as *A205/O101* (Barjas-Castro *et al.* 2000). The same study also reported that the *A201* allele behaved like an A^3 allele, instead of an A^2 allele, in a certain family.

A^3 alleles

A^3 alleles are responsible for the rare subgroup A_3 (Daniels, 1995). *A301* has a novel nonsynonymous substitution 871G>A (Asp291Asn) on the *A101* background (Yamamoto *et al.* 1993*d*). *A302* has two single base changes: the substitution 829G>A (Val277Met) characteristic of *O201* and the 1060delC typical of *A201* (Barjas-Castro *et al.* 2000). The same authors also described an interesting, but unresolved, example of an A^3 allele in a group A_3 individual:

four *O201*-defining base substitutions (646T>A, 681G>A, 771C>T and 829G>A) in homozygosity, and two *A201*-defining base changes (467C>T and 1060delC) and the *O101*-defining deletion (261delG) in heterozygosity. It is probable that this group A_3 individual is heterozygous for the *O202* allele (see Fig. 3) and a new hybrid A^3 allele in the form of *A201-O201-A201*.

A^x alleles.

A^x alleles are responsible for the rare subgroup A_x (Daniels, 1995). There are four A^x alleles reported to date (Fig. 1) Coincidentally, these four alleles have one to four base substitutions (646T>A, 681G>A, 771C>T and 829G>A) characteristic of *O201*. *Ax01* (also known as *A108* and *A^x(1)*) has 646T>A (Phe216Ile) on the *A101* background (Ogasawara *et al.* 1996*b*; Olsson & Chester, 1998; Yamamoto *et al.* 1993*c*). *Ax04* (or *A113*) has both 646T>A and 681G>A on the *A101* background, too (Ogasawara *et al.* 2001). Both *Ax02* and *Ax03* (also called *A^x(3)* and *A^x(2)*, respectively) have all four *O201*-defining substitutions, but *Ax02* has in addition the substitution 297A>G characteristic of the *B101* allele (Olsson & Chester, 1998). Since 297A>G is synonymous, the transferases encoded by these two alleles have identical amino acid sequence. It is very likely that the two amino acid substitutions (Phe216Ile and Val277Met) greatly reduce the enzymatic activity of the encoded A transferase and hence produce the weak phenotype A_x. It is noteworthy that none of the base substitutions found in the reported A^x alleles are novel.

R102 is identical to *Ax03* and *R103* to *Ax02* in the coding sequence analysed (exons 6 and 7), but the members of each matching pair differ in the sequence of intron 6 (Ogasawara *et al.* 2001; see below). Interestingly, the *R102* allele is expressed as group O in *R102/O* heterozygotes, but as group B(A) in *R102/B* heterozygotes (see below for explanation of the B(A) phenotype). This was demonstrated both in a small number of unrelated group O and group B(A) individuals, and also in a family in which the mother of genotype *R102/O* was typed as group O and a daughter of genotype of *R102/B* as group B(A). This serological discrepancy could probably be explained by allelic enhancement (Daniels, 1995). *R103* has been identified in some group O individuals. The *R103*-encoded transferase is predicted to be identical to the *R102*-encoded transferase in terms of their amino acid sequence. Thus, *R103* may also cause allelic enhancement although the *R103/B* genotype has yet to be found.

A^{el} alleles

A^{el} alleles are responsible for the rare subgroup A_{el} (Daniels, 1995). There are two A^{el} alleles reported in the literature (Fig. 1). *Ael01* (or *A109*) has a single G insertion in a series of seven G's at nt 798–804 (hereafter designated as 800insG) (Olsson *et al.* 1995; Ogasawara *et al.* 1996*b*). This novel insertion results in a complete alteration of the amino acid sequence after the glycine at position 268, immediately after the putative nucleotide-sugar binding site of the enzyme. The encoded transferase is expected to be 37 amino acids longer than the normal enzyme and 16 amino acids longer than the *A201*-encoded transferase. The significant reduction in the enzymatic activity is expected to result from the altered and lengthened C-terminal of the enzyme. *Ael01* accounts for the majority of A^{el} alleles.

Ael02 (or *A110*) possesses two *O201*-defining base substitutions (646T>A and 681G>A) on the *A102* background (Ogasawara *et al.* 1996*b*). In other words, the encoded enzyme has two amino acid substitutions (Pro156Leu and Phe216Ile). As in the case of A^x alleles mentioned above, it is presumably the amino acid substitution Phe216Ile that dramatically decreases the enzymatic activity of the encoded enzyme. Intriguingly, *Ael02* was also identified in five group O individuals (presumably of genotype *Ael02/O*) with weak serum anti-A antibody (Ogasawara *et al.* 2001). The authors argued that the A antigens on the red cells might be present in too trace an amount to be detected. The same authors also described an interesting family in which the father and the daughter both had the genotype *Ael02/O*, but their blood groups were

A_{el} and O lacking the serum anti-A antibody, respectively.

Heterogeneity of B alleles

Normal B alleles

There are also five *B* alleles that give normal blood group B with an *O* allele or with any one of themselves, and group A_1B with a normal A^1 allele: *B101* (the classical *B* allele), *B102*, *B103*, *B107* and *B108* (Fig. 2). These are the normal *B* alleles. As has been discussed above, the *B101* allele differs from the *A101* allele by seven single base substitutions (three synonymous and four non-synonymous). The other four *B* alleles each differ from *B101* at a single position: nt 930 for *B102*, nt 657 for *B103*, nt 526 for *B107* and nt 297 for *B108* (Ogasawara *et al.* 1996*a*, 1998, 2001). Interestingly, these are also four of the seven positions where *B101* and *A101* differ, and these four alleles have the *A101* sequence at these positions. Three substitutions are synonymous, but one (nt 526) is nonsynonymous. Since the amino acid substitution corresponding to nt 526 does not affect the substrate specificity of the encoded transferase (Yamamoto & Hakomori, 1990), it is understandable that the *B107*-encoded transferase also behaves like the *B101*-encoded transferase. Except *B101*, all alleles are rare (< 1 %) and have been reported to date in Japanese only (Ogasawara *et al.* 1996*a*, 1998, 2001).

B^3, B^x and B^{el} alleles

These alleles are responsible for the rare subgroups B_3, B_x and B_{el} respectively (Daniels, 1995). There are two B^3, one B^x and two B^{el} alleles reported to date in the literature (Fig. 2). One B^3, one B^x and two B^{el} alleles each have one additional nonsynonymous substitution on the *B101* haplotype background: *B301* with 1054C>T (Arg352Trp), *Bx01* with 871G>A (Asp291Asn), *Bel01* with 641T>G (Met214Arg) and *Bel02* with 669G>T (Glu223Asp) (Yamamoto *et al.* 1993*d*; Ogasawara *et al.* 1996*b*). Presumably, these amino acid substitutions reduce the enzymatic activities of the encoded B transferase. All substitutions are new except Arg352Trp which is also found in another rare allele, *A202* (see Fig. 1). The other B^3 allele, *B302*, differs from *B101* by two base substitutions: 646T>A (Phe216Ile) characteristic of the *O201* allele and the same sequence at nt 657 as the *O201* (or *A101*) allele (Ogasawara *et al.* 2001). Thus, *B302* has the *O201* sequence at these two positions.

B(A) alleles

The group B(A) individuals have normal B antigens on red cells with anti-A in the serum, but their red cells are weakly agglutinated by some highly potent monoclonal anti-A reagents (Daniels, 1995). The first reported *B(A)* allele, *B(A)01*, differs from the *B101* allele at two positions, nt 657 and 703 (Fig. 2; Yamamoto *et al.* 1993*c*). At these positions, *B(A)01* has the same sequence as the *A101* allele. In other words, the encoded enzyme has the composition of BABB for the four characteristic amino acid residues (176, 235, 266 and 268) that distinguish the B transferase from the A transferase. This result was unexpected because the transferase with the composition of BABB showed reactivity only with anti-B reagents in the transfection experiments using HeLa cells (Yamamoto & Hakomori, 1990). But it is also possible that the experimental system using HeLa cells is not sensitive enough to detect the activity. The second *B(A)* allele, *B(A)02* (or $B(A)^{700}$), is due to a novel nonsynonymous substitution 700C>G on the *B101* background, which produces the amino acid substitution Pro234Ala (Yu *et al.* 1999). It is intriguing to find that this substitution is just ahead of the second (residue 235) of the four characteristic amino acid residues just mentioned above. It is also notable that alanine has a smaller side chain than proline, and that the nucleotide-sugar donor UDP-GalNAc for A transferase is larger than the UDP-Gal for B transferase. It may not be coincidental at all that both *B(A)*-encoded enzymes involve critical amino acids adjacent to each other (residues 234

and 235) on a typical B transferase polypeptide backbone.

Cis-AB alleles

Cis-AB is a rare ABO blood group characterized by the presence of A, weakened B and elevated H antigens on the red cells with anti-B in the serum non-reactive against autologous red cells (Daniels, 1995). It is thus serologically distinct from the classical AB and the B(A) phenotypes. It is due to the inheritance of a single chromosome encoding an enzyme with both A and B transferase activities. *Cis-AB01* (or *C101*) has the substitution 803G>C (Gly268Ala) on the *A102* background (Fig. 2; Fukumori *et al.* 1996*b*; Hosoi *et al.* 1993; Ogasawara *et al.* 1996*b*; Yamamoto *et al.* 1993*a*). The alanine at position 268 is specific to the B transferase. Thus, the encoded enzyme has the composition of AAAB for the four characteristic amino acid residues and thus chimeric A/B transferase activities, which were correctly predicted by the transfection experiments using cell lines (Yamamoto & Hakomori, 1990; Yamamoto & McNeill, 1996). The first series of transfection experiments also demonstrated that a transferase with the composition BBAB at the four characteristic amino acid residues exhibited chimeric A/B transferase activities (Yamamoto & Hakomori, 1990). This is the predicted enzyme encoded by the second reported *cis-AB* allele, namely, *cis-AB02* or *cis-AB*^{var} (Midsuf *et al.* 2000). Within the limited region studied, *cis-AB02* differs from *B101* at nt 796 where it has the same sequence as the *A101* allele.

Heterogeneity of O alleles

There are nearly forty different *O* alleles reported to date that do not produce catalytically active transferase (Fig. 3). They are also known as null alleles or non-functional alleles.

Alleles O101 and O201

Two *O* alleles are common in all populations studied so far: *O101* and *O201* (Fukumori *et al.* 1996*a*; Kang *et al.* 1997; Nishimukai *et al.* 1996; Ogasawara *et al.* 1996*a*, 2001; Olsson & Chester, 1996*a*; Olsson *et al.* 1998; Roubinet *et al.* 2001; Watanabe *et al.* 1997; Yip, 2000; Yip *et al.* 1995). *O101* has a single G deletion at nt 261 (261delG) when compared to the *A101* cDNA sequence (see above; Yamamoto *et al.* 1990*a*). *O201*, also known as *O*^{1v} or *O02*, differs from *O101* by nine base substitutions at nt 106, 188, 189, 220, 297, 646, 681, 771 and 829 (Yamamoto *et al.* 1990*a*; Olsson & Chester, 1996*a*). Although three of these are nonsynonymous and produce amino acid substitutions (Val36Phe, Arg63His and Pro74Ser), the same deletion (261delG) that renders *O101* non-functional produces the same effect on *O201*. For the sake of convenience, all other *O* alleles will be described as *O101*-like, *O201*-like or non-deletional alleles. The division between *O101*-like and *O201*-like is sometimes blurred and arbitrary because some alleles are hybrids of both *O101* and *O201*.

O101-like alleles

There are eighteen *O101*-like alleles that are characterized by the presence of the 261delG and at least one additional point mutation (Fig. 3). They can be grouped into three categories. In the first category, there are six *O101*-like alleles that each possess one additional novel base substitution not found in any of the common *ABO* alleles: *O106* (or *O*$^{1(G\,190A)}$) with 190G>A, *O115* (or *Ovar.tlse08*) with 318C>T, *O104* with 454T>C, *O102* with 579T>C, *O117* (or *Ovar.tlse02*) with 768C>A and *O118* (or *Ovar.tlse07*) with 926A>G (Ogasawara *et al.* 1996*a*, 1998; Olsson *et al.* 1997; Roubinet *et al.* 2001). In other words, these new point mutations occurred on the *O101* haplotype background. In the second category, there are 10 *O101*-like alleles that have one or more base substitutions characteristic of the common *ABO* alleles. In a sense, they can be called hybrid alleles without implying the mechanism of their occurrence. *O107* (or *O*$^{1(C\,467T)}$) has the substitution 467C>T characteristic of the *A102* allele (Olsson *et al.* 1997). *O110* (or *O*^{v3}) has the substitution 657C>T characteristic of the *B101* allele (Yip, 2000). Three alleles (*O111*, *O108* and *O112*) have one or more point mutations (1060delC and 467C>T)

that are found in the *A201* allele (Barjas-Castro *et al.* 2000; Olsson *et al.* 1997). *O112* also has the substitution 646T>A typical of the *O201* allele. Five alleles (*O105*, *O113*, *O114*, *O119* and *O202*) each have 1–4 base substitutions characteristic of the *O201* allele (Barjas-Castro *et al.* 2000; Olsson *et al.* 1997; Roubinet *et al.* 2001). These substitutions include 188G>A, 189C>T, 646T>A, 681G>A, 771C>T and 829G>A. Some authors preferred to describe *O119* (or *Ovar.tlse05*) and *O202* (or O^{v1}) as *O201*-like because these two alleles contain three and four substitutions defining the *O201* allele, respectively. In the third category, there are two alleles with the substitution 467C>T characteristic of the *A102* allele and 1–2 novel substitutions: *O109* (or O^{v2}) with 318C>T and *O116* (or *Ovar.tlse04*) with 318C>T and 729C>T (Roubinet *et al.* 2001; Yip, 2000).

In most instances, these *O101*-like alleles are quite rare with only a few examples recorded for the ethnic groups under study and sometimes in one population only. Interestingly, though found at very low frequencies in several populations, *O109* accounts for 19% of all the *O* alleles in Akans from Ivory Coast, who are known to have high frequencies of *O* alleles. It thus seems likely that, for the *O116* allele reported at low frequency in Akans only, the substitution 729C>T occurred on the common *O109* haplotype background.

O201-like alleles

There are thirteen *O201*-like alleles that are characterized by the presence of both 261delG and 297A>G, and in most cases a few other substitutions defining the *O201* allele such as 646T>A, 681G>A, 771C>T and 829G>A (Fig. 3). Like *O101*-like alleles, they can also be grouped into three categories. In the first category, there are six alleles that each have one additional new base substitution not found in any other common *ABO* allele and on a typical *O201* haplotype background for the region (exons 6 and 7) under study. They are *O212* (or *Ovar.tlse01*) with 351G>A, *O213* (or *Ovar.tlse11*) with 498C>T, *O214* (or *Ovar.tlse12*) with 538C>T, *O209* (or O^{v5}) with 542G>A, *O210* (or O^{v6}) with 595C>T and *O203* with 721C>T (Ogasawara *et al.* 1996*a*; Olsson *et al.* 1998; Roubinet *et al.* 2001; Yip, 2000). Thus, these six alleles were derived from the *O210* allele. In the second category, there are five hybrid alleles that have one or more base substitutions characteristic of the common *ABO* alleles. *O103* has the same sequence as the *A101* (or *A102* or *O101*) allele at the four characteristic positions (nt 646, 681, 771 and 829) (Ogasawara *et al.* 1996*a*). *O208* (or O^{v7}) has the same sequence as the *A101* allele at nt 681 (Yip, 2000). *O205* and *O206* (or O^{1}-O^{w}*(1)* and O^{1}-O^{w}*(2)*, respectively) have the same sequence as the *A101* allele at nt 106, 188 and 189, and the same situation also applies to nt 220 for *O205* (Olsson *et al.* 1997). *O207* (or O^{w}-*B*) is a hybrid of alleles *O201* and *B* with the same sequence as the *B* allele at seven characteristic positions (nt 526, 657, 703, 796, 803, 930 and 1096) (Olsson *et al.* 1997; Roubinet *et al.* 2001). In the third category, there are two alleles each with a new base substitution on the *O208* haplotype background: *O211* (or *Ovar.tlse03*) with 529G>A and *O204* (or $O^{w,681\&1054}$) with 1054C>T (Olsson *et al.* 1997; Roubinet *et al.* 2001).

O208 is found at low frequencies in several populations and, interestingly, its frequency increases from north to south for Basques (from France and Spain, 0.5% of all *O* alleles), Berbers (from Morocco, 9%) and Akans (from Ivory Coast, 19%) (Roubinet *et al.* 2001; Yip, 2000). *O209* accounts for 4% of all *O* alleles in Cayapas from Ecuador, 12% in Aymaras from Bolivia (Roubinet *et al.* 2001), and interestingly 43% in Amerindians from Brazil (Olsson *et al.* 1998). Other alleles are found at low frequencies in the populations under study. One important point emerging from these studies is that a global study of sequence polymorphism in various populations should be undertaken before any given allele is considered as specific to one population.

Non-deletional O alleles

There are now four examples of *O* alleles that do not possess the 261delG characterizing the

O101 and *O201* alleles (the so-called non-deletional *O* alleles). They are *O303*, *O304*, *O301* and *O302* (Fig. 3). *O303*, also known as O^2 or *O03*, differs from *A101* by five single base substitutions within the coding sequence at nt 53, 220, 297, 526 and 802 (Yamamoto *et al.* 1993*b*; Amado *et al.* 2000). Four of these base substitutions (53G>T, 220C>T, 526C>G and 802G>A) result in amino acid substitutions (Arg18Leu, Pro74Ser, Arg176Gly and Gly268Arg, respectively). The substitution 220C>T is characteristic of the *O201* allele, while 297A>G and 526C>G are typically found in the *B101* allele. The other two are novel mutations. Note that the novel mutation 802G>A changes the amino acid at the same position 268 as the substitution 803G>C characteristic of the *B101* allele, but to a different one (arginine, instead of alanine in the B transferase). This non-deletional allele has been shown to be non-functional by the lack of immuno-detectable surface A and B antigens on HeLa cells (Yamamoto *et al.* 1993*b*) and the lack of A transferase activity in Sf9 cells and the culture supernatant (Amado *et al.* 2000) after transfection of the cells with appropriate *O303* cDNA constructs. However, the intact reading frame of this allele does produce a protein recognized by monoclonal antibodies against the A/B transferase, as has been shown by immunohistochemistry in *O303*-transfected Sf9 cells and gastric carcinomas genotyped as *O101/O303* (Amado *et al.* 2000). It seems likely that the non-conservative amino acid substitution Gly268Arg inactivates the *O303*-encoded transferase because this amino acid residue is in the putative nucleotide-sugar binding site. This hypothesis is consistent with the results of the early transfection and expression experiments conducted by Yamamoto & Hakomori (1990). The *O303* allele has not yet been reported in Korean, Japanese and Chinese populations (Fukumori *et al.* 1996*a*; Kang *et al.* 1997; Ogasawara *et al.* 1996*a*; Yip, 2000), suggesting that this allele may be extremely rare or even does not exist in Oriental populations. Interestingly, this allele was not found in Amerindians either (Franco *et al.* 1995; Olsson *et al.* 1998). However, it is found at low frequencies (0.7% –2.8%) in various Caucasian populations (Gassner *et al.* 1996; Nishimukai *et al.* 1996; Olsson & Chester, 1995; Pearson & Hessner, 1998; Watanabe *et al.* 1997; Yip, 2000). It is also present at low frequency in blacks (Franco *et al.* 1995).

The second non-deletional *O* allele (*O304*) has the insertion 800insG on the *A201* haplotype background (Fig . 3; Olsson & Chester, 1996*c*). It is also known as O^3 or *O08*. The co-existence of 800insG and 1060delC means that the encoded protein has the same length as the A or B transferase but with a completely different amino acid sequence after the glycine at position 268 (due to 800insG), immediately after the postulated nucleotide-sugar binding site of the enzyme. This altered amino acid sequence presumably inactivates the enzyme. Interesting is the observation that the allele *Ael01* with 800insG alone produces an enzyme with an altered and lengthened C-terminal which dramatically reduces the enzymatic activity, instead of completely inactivating the enzyme as in the case of the *O304* allele. The last two non-deletional *O* alleles are due to novel mutations near the 3′ end of the gene. *O301* has the non-synonymous substitution 893C>T (Ala298Val) on the *A102* background while *O302* has the nonsense mutation 927C>A (Tyr309Stop) on the *A101* background (Ogasawara *et al.* 2001). All these three non-deletional *O* alleles are very rare.

Critical amino acid positions and residues

The amino acid differences of the transferases for the subgroups are summarized in Table 1. It is well established that the amounts of A or B antigens on red cells or their reactivity with the corresponding antibodies are in the following decreasing order: $A_1 > A_2 > A_3 > A_x > A_{el}$ and $B > B_3 > B_x > B_{el}$ (Daniels, 1995). Thus, the most critical amino acids for the enzymatic activity are found at positions 214 (Met>Arg), 216 (Phe>Ile), 223 (Glu>Asp) and 291 (Asp>Asn). The amino acid substitutions at positions 277 (Val>Met), 337 (Arg>Gly) and 352 (Arg>Gly/

Table 1. *Critical amino acid positions and residues of transferases encoded by the ABO gene*

(Blood group) Phenotype	Amino acid position and residue*											
	156	176	214	216	223	268	277	291	298	337	352	C-terminal
A_1	Pro	Arg	Met	Phe	Glu	Gly	Val	Asp	Ala	Arg	Arg	
A_1	Leu	—	—	—	—	—	—	—	—	—	—	
A_2	Leu	—	—	—	—	—	—	—	—	—	—	Longer†
A_2	—	—	—	—	—	—	—	—	—	—	—	Longer
A_2	—	—	—	—	—	—	—	—	—	—	Gly	
A_2	—	—	—	—	—	—	—	—	—	—	Trp	
A_2	Leu	—	—	—	—	—	—	—	—	Gly	—	
A_3	—	—	—	—	—	—	—	Asn	—	—	—	
A_3	—	—	—	—	—	—	Met	—	—	—	—	Longer
A_x	—	—	—	Ile	—	—	Met	—	—	—	—	
A_x	—	—	—	Ile	—	—	—	—	—	—	—	
A_{el}	Leu	—	—	Ile	—	—	—	—	—	—	—	
B	—	Gly	—	—	—	Ala	—	—	—	—	—	
B_3	—	—	—	—	—	—	—	—	—	—	Trp	
B_3	—	—	—	Ile	—	—	—	—	—	—	—	
B_x	—	—	—	—	—	—	—	Asn	—	—	—	
B_{el}	—	—	—	—	Asp	—	—	—	—	—	—	
B_{el}	—	—	Arg	—	—	—	—	—	—	—	—	
O	Leu	—	—	—	—	—	—	—	Val	—	—	
O	—	Gly	—	—	—	Arg	—	—	—	—	—	

* All amino acid substitutions are due to novel mutations except Pro156Leu, Arg176Gly, Phe216Ile and Gly268Ala.

† Longer C-terminal domain of the transferase refers to that arising from the deletion 1060delC.

Trp) may also lead to a decrease in the enzymatic activity. But the amino acid substitution at position 156 (Pro > Leu) most probably has little effect. In addition, the amino acid substitutions at positions 268 and 298 are definitely crucial to the enzymatic activity since Gly268Arg and Ala298Val completely inactivate the enzyme.

extensive sequence variation in the non-coding region

Extensive sequence variation in the coding region of the *ABO* gene produces a large number of different *ABO* alleles, as has been discussed in detail above. Likewise, extensive sequence variation in the non-coding region of the gene has also been reported: variation in minisatellite repeats in the 5′ untranslated region (UTR), 21 single nucleotide polymorphisms (SNPs) in intron 6 and another SNP in the 3′ UTR. The patterns of these variations show specific relationship with individual major *ABO* alleles. The discovery and genotyping of SNPs in intron 6 are crucial to the elucidation of the origins of some novel alleles because the haplotypes of these SNPs serve as footprints of past recombination events occurring within the *ABO* locus.

Variation in minisatellite repeats in 5′ UTR

The transcription of the *ABO* gene is dependent on the binding of the transcription factor CBF/NF-Y to an enchancer-active minisatellite located about −3.7 kb upstream of the transcription start site (Kominato *et al.* 1997), as has been discussed above. The basic building block of this minisatellite is a 43 bp repeat unit. There is variation in the copy number of the minisatellite repeats with a correlation between the *ABO* alleles and the copy numbers: alleles *A101*, *A102* and *O303* have only one repeat unit while alleles *A201*, *B101*, *O101* and *O201* have four tandem repeat units (Irshaid *et al.* 1999). This allele-related variation in minisatellite repeats is also observed in various geographical areas and ethnic groups with very few excep-

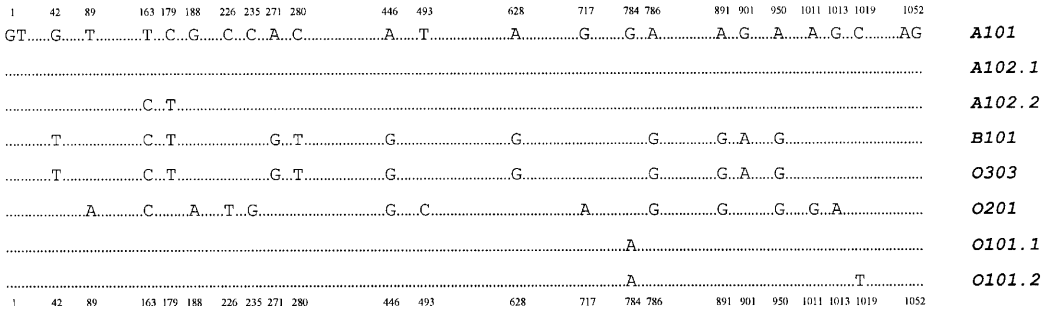

Fig. 4. Schematic comparison of the nucleotide sequence for intron 6 of the *ABO* gene. The *A101* allele is used as the reference for comparison and only differences from this sequence are indicated.

tions. In addition, all alleles with only one repeat unit have a nucleotide substitution (G>A) at nt 41 of the 43 bp repeat unit (Irshaid *et al.* 1999; Yu *et al.* 2000). Furthermore, 35% of *O101* alleles and all *O201* alleles have another base substitution (G>C) also at nt 41, but only in the first of the four repeat units (Yu *et al.* 2000). It is noteworthy that there are so far no intermediate forms of the minisatellite (i.e. 2–3 repeats) reported.

The relationship between the copy number of the repeat units and their transcriptional activities was studied by transient transfection of gastric cancer cell line KATO III with reporter plasmids containing these enhancer structures (Yu *et al.* 2000). The transcriptional activity of the A^1-gene enchancer (one repeat unit) was found to be less than 1% of that of the *B*-gene enhancer (four repeat units). The difference between the transcriptional activities of these two enhancers became more significant when acting in concert with the native promoter of the *ABO* gene. Moreover, the transcriptional activity was found to be elevated with an increase in the number of repeat units. This situation is not unique to the *ABO* gene. For example, the insulin-linked polymorphic region with a longer minisatellite also shows greater transcriptional activity than that with a shorter minisatellite (Kennedy *et al.* 1995). However, it should be noted that these *in vitro* assays may not truly reflect *in vivo* status. It also seems likely that there are other regulatory elements not yet identified but affecting the expression of the *ABO* gene. This conjecture can be made at least in light of the fact that the number of antigen sites on red cells is, on the contrary, slightly greater for A_1 than for B individuals (Daniels, 1995).

SNPs in intron 6

Extensive sequence variation in intron 6 was first discovered by Suzuki *et al.* (1997), and later confirmed and extended by Olsson & Chester (1998) and Ogasawara *et al.* (2001). Intron 6 of the *ABO* gene is 1052 bp long. There are 21 SNPs within this short distance with an average of 20 SNPs per kb. Thus, the SNP density in intron 6 is about 40-fold greater than the average global density of 0.5 SNP per kb (The International SNP Map Working Group, 2001). Not surprisingly, the haplotypes of these SNPs are limited in number and related to the major *ABO* alleles (Fig. 4). Comparison is again made against the *A101* allele as the reference. In the majority of cases (65%), *A102* has a sequence identical to that of *A101* in intron 6 (designated as *A102.1*). The rest of *A102* (designated as *A102.2*) has nucleotide substitutions at nt163 (T>C) and 179 (C>T). Intriguingly, these two nucleotides are identical to those of *B101* and *O303*, indicating the possible occurrence of a past gene conversion event. *B101* and *O303* are identical in intron 6 and differ from *A101* by eleven base substitutions. *O201* differs from *A101* by 13 base substitutions, five of which are shared with the intron sequence of alleles *B101* and *O303*. Most cases of *O101* (designated as *O101.1*) differ from *A101* in intron 6 just by one substitution (784G>A). In about 7% of cases, *O101* has

another substitution (1019C>T) in addition to 784G>A (designated as *O101.2*). The discovery of these *ABO* allele-specific haplotypes of intron 6 SNPs is very significant because they can mark the breakpoints of recombination events occurring within the *ABO* locus, as will be discussed below.

Sequence variation in 3′ UTR

There is one base substitution found after the termination codon (nt 1063–1065) at nt 1096 (Olsson & Chester, 1995). This novel substitution (1096G>A) is present only in *B101* and *O303* alleles. The discovery of this SNP allows the design of a single-tube genotyping method of the *ABO* locus based on multiplex PCR followed by restriction analysis. This method allows the distinction of five major alleles (*A101*, *A201*, *B101*, *O101* and *O303*).

Role of point mutatons and recombination in the genetic diversity of the *ABO* locus

The number of *ABO* alleles shown in Figures 1–3 totals 71. It is expected that new alleles will continuously be reported in the future. For the sake of easy discussion, the following will be regarded as common alleles: *A101*, *A102*, *A201*, *B101*, *O101*, *O201* and *O303*. The fact that a few other *O* alleles are common in some populations (Roubinet *et al.* 2001) but are not regarded as common alleles in the following discussion, does not negate the arguments.

Role of point mutations

There is a total of 45 mutation sites in the coding sequence, mostly in exons 6 and 7, of the *ABO* locus among these 71 alleles. However, there are a total of 46 point mutations, because one particular mutation site harbours two different mutations (1054C>T and 1054C>G). Of these 46 point mutations, 21 are found among the seven common alleles with 13 being transitions, six transversions and two single base deletions. On the other hand, there are 25 new mutations not found in these common alleles with 18 being transitions, six transversions and one single base insertion. In summary, there are 31 transitions, 12 transversions, two deletions and one insertion. Transitions make up 72% (31/43) of all base substitutions. Transitions occurring in the CpG dinucleotides (i.e. CG>TG or CG>CA) comprise 37% (16/43) of all base substitutions or 52% (16/31) of all transitions. Such transitions are explicable by methylation-mediated deamination of 5-methylcytosine found in CpG dinucleotides (Cooper & Krawczak, 1993). These proportions are also in good agreement with those obtained for a much larger compilation (Cooper & Krawczak, 1993).

Excluding the seven common alleles, there are 64 alleles reported to date. Of these, 30 alleles (47%) contain new mutations. A few of these mutations are found in more than one allele: 1054C>T and 318C>T each in three alleles, and 800insG and 871G>T each in two alleles (Fig. 1 to 3). The other 34 alleles are due to different combinations of mutations that are found in the common alleles. In other words, more than half of these *ABO* alleles (53% or 34/64) are not accounted for by new mutations. As will be shown below, the occurrence of these remaining *ABO* alleles can very probably be accounted for by reciprocal recombination or gene conversion.

Role of reciprocal recombination

Sequence analysis of newly identified alleles first led Ogasawara *et al.* (1996*b*) to propose that alleles *A204* and *Ael02* (see Fig. 1) might have originated by recombination between *B101* and *O201*, and between *A102* and *O201*, respectively. Sequence analysis of the last two exons and intron 6 allowed Suzuki *et al.* (1997) to document unambiguously a *de novo* recombination within the *ABO* locus to produce a hybrid allele in the child in a paternity case where the putative father was only excluded by ABO blood grouping (mother, B; child, A_1; and putative father, O) among many polymorphic markers tested. Recently, detailed haplotype analysis of the SNPs in intron 6 has allowed many more past re-

combination events to be recorded, because the haplotypes of these intronic SNPs are *ABO* allele-specific (see above). In fact, a sizable proportion of the *ABO* alleles not accounted for by new mutations are the products of reciprocal recombination or crossover (Fig. 5). It is extremely difficult, if not impossible, to demonstrate simultaneously the two reciprocal recombinant chromosomes or alleles generated from a single crossover event in humans. But, the reciprocal recombinant alleles generated from different crossover events involving the same pair of parental alleles can sometimes be identified in unrelated individuals, as will be shown below.

Reciprocal recombination between *A102* and *B101* can generate *B108* as the *A102-B101* product, and *A112*, *A104* and *B102* as the *B101-A102* product (Fig. 5). The recombination site is in intron 6 at nt 280–446 in one example of *B108* and at nt 786–891 in one case of *A112* (Ogasawara *et al.* 2001). The recombination site is in the junction of intron 6 at nt 930 and exon 7 at 526 for twelve examples of *A104* in unrelated Japanese (Ogasawara *et al.* 2001; Suzuki *et al.* 1997). Interestingly, the recombination site is the same for another example of *A104* in a Japanese child resulting from a *de novo* recombination (Suzuki *et al.* 1997). In other words, the recombination generating *A104* has occurred at least twice in the past. The recombination site for *B102* is in exon 7 at nt 803–930. Instead of *A102*, *A101* can also be one of the parental alleles in forming *B108*, *A104* and *B102*. Similarly, *O101* can replace *A102* as one of the parental alleles in generating *A104* and *B102*. The *A104* allele was in fact reported by Suzuki *et al.* (1997) for the case (designated *b-o1*) arising from a *de novo* recombination (see above).

Reciprocal recombination between *B101* and *O201* can produce *Ax02* and *A204* as the *B101-O201* product, and *O207* as the *O201-B101* product. Several examples of *Ax02* have been described: $A^x(3)$ in a Swedish family (Olsson & Chester, 1998), *R103.1* in one Japanese, and *R103.2* in two Japanese individuals (Ogasawara *et al.* 2001). $A^x(3)$ and *R103.1* have the same recombination site in intron 6 at nt 717–901 while *R103.2* has the recombination site in the junction of intron 6 at 1013 and exon 7 at 526. The recombination site for *A204* (designated as *R101* and *b-o1v* by different authors) is in exon 7 at nt 703–771 (Ogasawara *et al.* 1996*b*, 1998; Suzuki *et al.* 1997). On the other hand, the recombination site for *O205* (or O^{w}-*B*) is somewhere between exon 6 at 261 and exon 7 at 526 (Olsson *et al.* 1997; Roubinet *et al.* 2001), and has not been defined more precisely because its intron 6 has not been sequenced and analysed in detail. Note that *A207* is found in both Swedish and Akans. In summary, recombination between *B101* and *O201* has occurred at least five times in the past.

Reciprocal recombination between *O101* and *O201* can create *O202* as the *O101-O201* product and *O103* as the *O201-O101* product. *O202* (designated *o1-o1v* by Suzuki *et al.*) has a recombination site in intron 6 at nt 235–446 for eight cases in Japanese (Ogasawara *et al.* 2001; Suzuki *et al.* 1997). On the other hand, the recombination site for *O103* is in intron 6 at nt 950–1011 in three examples in Japanese (Ogasawara *et al.* 2001), and in the junction of intron 6 at nt 1013 and exon 7 at nt 646 for one example (named *o1v-o1* or *o1v-a1*) found in an Udeghe-Russian admixed individual (Suzuki *et al.* 1997). Note that *A101* can replace *O101* as one of the parental alleles in generating *O103*. By the same token, *O205* and *O206* (also known as O^{1}-$O^{w}(1)$ and O^{1}-$O^{w}(2)$, respectively; see Fig. 3) could also be the products of recombination between *O201* and *O101* (Olsson & Chester, 1997) or *A101* occurring probably around intron 5 and intron 4, respectively.

Reciprocal recombination between *O201* and *A101* can lead to the formation of *O103* as the *O201-A101* product (see above) and *Ax03* as the *A101-O201* product. The recombination site for *Ax03* is in intron 6 at nt 188–226 in a Swedish family (named $A^x(2)$; Olsson & Chester, 1998), and at nt 959–1011 for five examples (named as *R102*) in Japanese (Ogasawara *et al.* 2001). Note that *R102* and *O103* have the same recombination junction region and thus form a match-

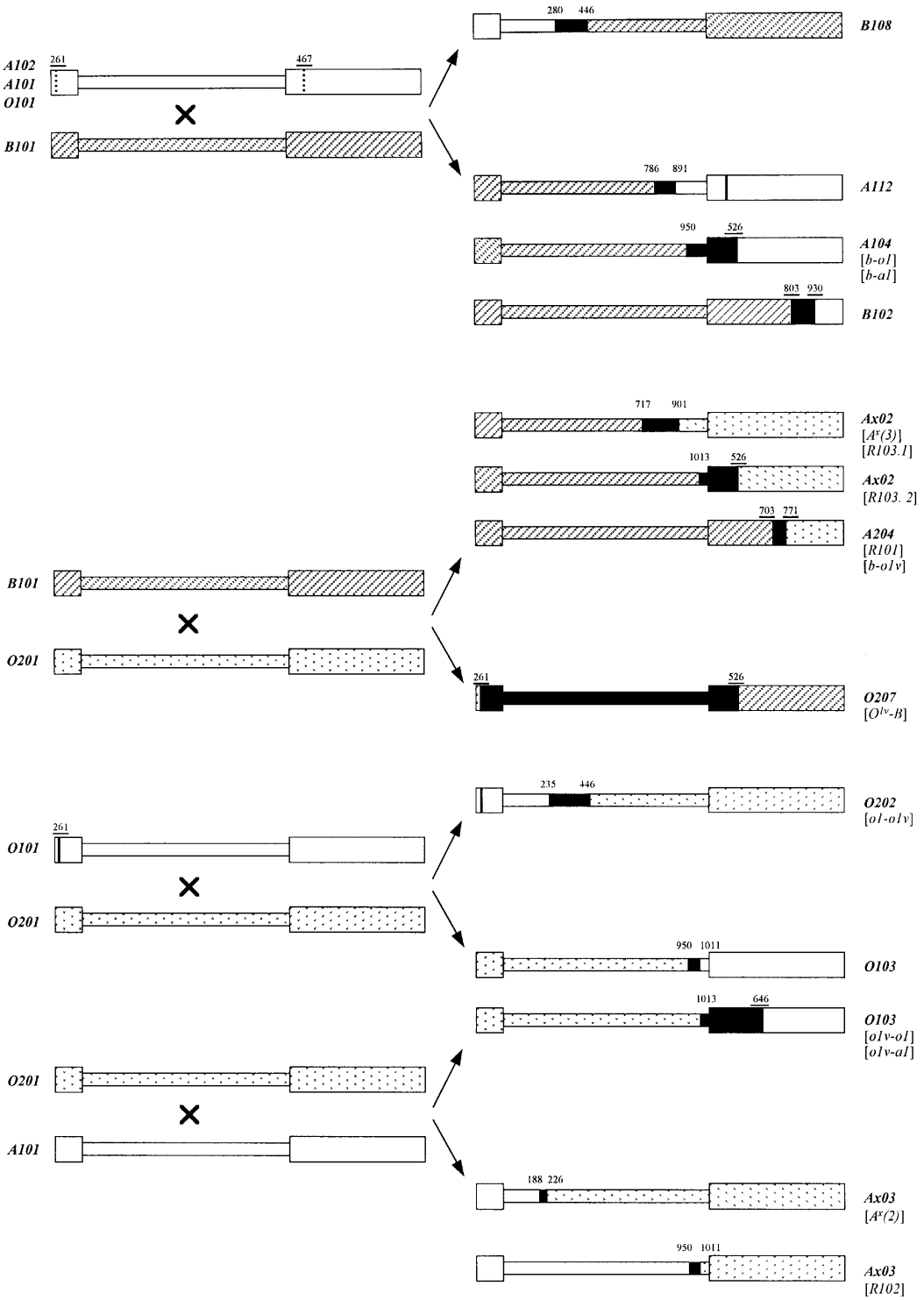

Fig. 5. Generation of new *ABO* alleles through reciprocal recombination between common alleles. Only exons 6 and 7 and intron 6 (the narrow bar in the middle) are shown. The parental alleles of recombination are shown on the left and the recombinant products on the right. The recombination junction regions are indicated as dark boxes with the nucleotide positions indicated. Underlined nucleotide positions refer to those in exons. Note that the nucleotide sequences of intron 6 for *B102*, *A204* and *O207* have not yet been examined.

ing pair of reciprocal recombinant products. Obviously, *A102.1* can replace *A101* as one of the parental alleles in producing *Ax03*.

In summary, the recombination sites are in intron 6 and spread into exon 7 (Fig. 5). In particular, the breakpoints tend to cluster around nt 235–446, nt 717–901and nt 950 of intron 6 to nt 526 of exon 7. Whether these are hotspots of recombination within the *ABO* locus remains to be determined.

Reciprocal recombination (or crossover) within the *ABO* locus has been invoked to explain the occurrence of the above-mentioned hybrid alleles. First, the recombination junction region can be defined quite precisely for most of these hybrid alleles through haplotype analysis of the SNPs in intron 6. Second, the seemingly reciprocal recombinant alleles can be demonstrated for different pairs of parental alleles. Third, there are two perfect chi sequences (GCTGGTGG) found within the *ABO* locus: one in intron 6 at nt 853–860, and another in intron 3 at nt 269–276 (numbering within the intron) found in opposite orientation in the anti-sense strand. This is intriguing because the chi sequence has been considered to be associated with recombination (Smith, 1994). Finally, the *de novo* recombination documented by Suzuki *et al.* (1997) serves as a very good example showing the existence of such a process in generating new *ABO* alleles. Accumulation of point mutations is much less likely a mechanism of producing these hybrid alleles because this would involve many mutations occurring in succession, particularly those in intron 6. On the other hand, the non-reciprocal transfer of a DNA segment from a donor allele to an acceptor allele (i.e. gene conversion) cannot be completely ruled out, though unlikely, as a mechanism of generating the above-mentioned hybrid alleles. This stems from the fact that genetic markers flanking the *ABO* locus have not been examined to show whether the recombination events were reciprocal (i.e. crossovers) or non-reciprocal (i.e. gene conversion). However, one study did manage to document that the 3′ ends of the hybrid alleles *A^x(2)* (an *A101-O201* product) and *A^x(3)* (a *B101-O201* product) comprised *O201* sequence even about 1.35 kb downstream the *ABO* coding sequence (Olsson & Chester, 1998).

There are other rare alleles which might have been generated by recombination: *O107*, *O111* and *O108* (Fig. 3). Recombination between *O101* and *A102* could have produced *O107*, although the mutation 467C>T (characteristic of *A102*) occurring recurrently on the *O101* background is also possible. *O111* and *O108* could also have been produced by recombination between *O101* and *A201* at different recombination sites. This speculation is however very difficult to prove because the intron 6 haplotypes of the parental alleles involved are either identical or almost identical.

A close examination of Figure 5 reveals that some recombinant alleles show a completely different specificity from their parental alleles involved in the recombination events: *A104* from *O101*×*B101*, *Ax02* and *A204* from *O201*×*B101*. If the recombination occurs *de novo*, then parent–child discrepancy in the ABO blood group may occur. This is best illustrated by the *A204* allele identified in the child in a paternity case reported by Suzuki *et al.* (1997). In fact, hybrid alleles resulting from recombination might explain the early paradoxical observation that A_x progeny can be produced from group O parents (van Loghem & van der Hart, 1954; Beckers *et al.* 1955). An *O303-O201* hybrid allele could be created by reciprocal recombination in a group O parent of the genotype *O201/O303*. This hybrid allele would contain exon 6 from *O303* and exon 7 from *O202*, and would be similar to the *Ax02* allele mentioned above because the inactivating 261delG characteristic of *O101* and *O201* is not found in exon 6 of *O303*. By the same token, an *O303-O101* hybrid allele from a group O parent of the genotype *O101/O303* would be expected to behave like an *A*^{1} allele.

Role of gene conversion

There are many alleles that contain one or two base substitutions characteristic of the common alleles and found on different haplotype back-

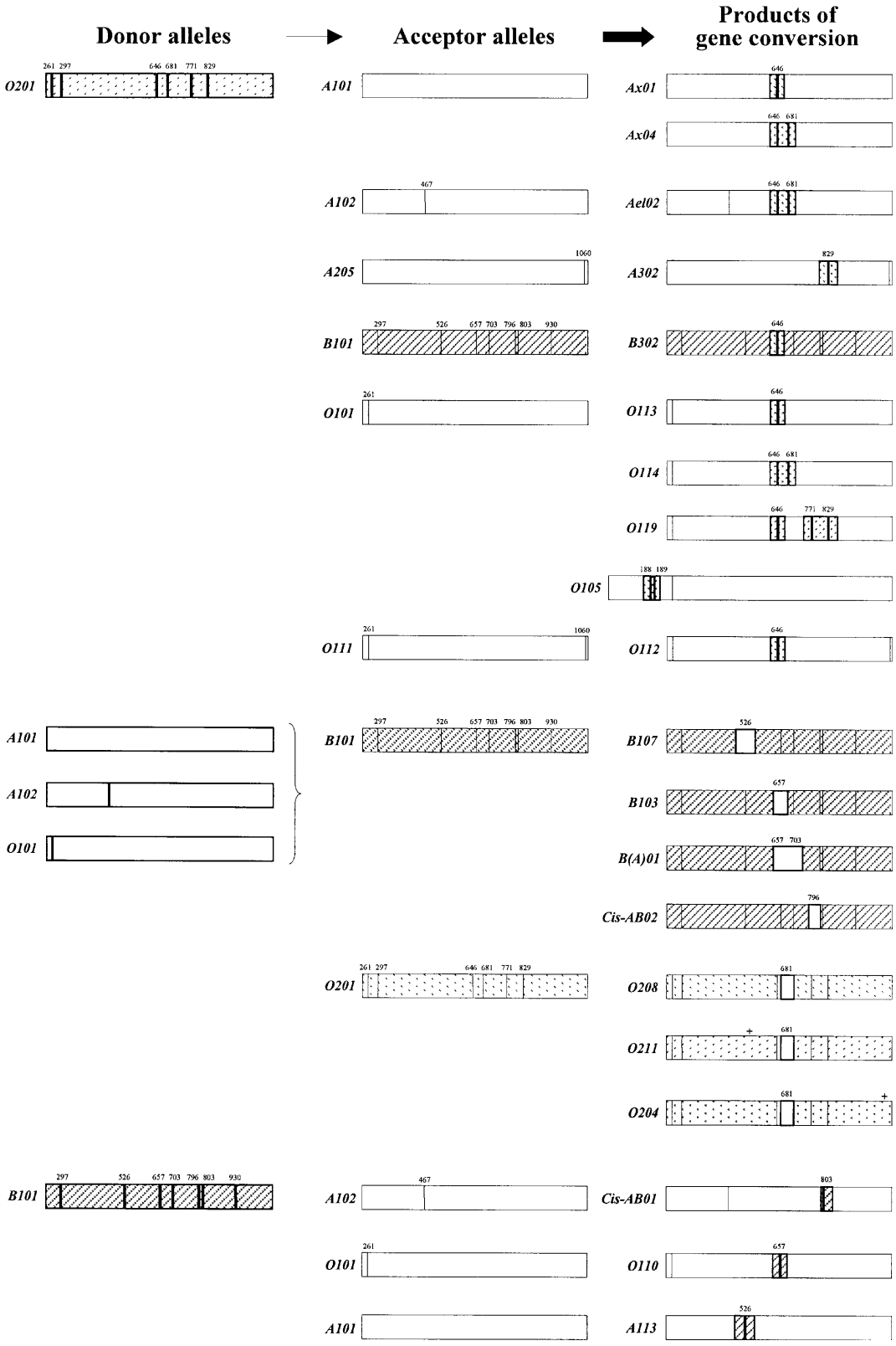

Fig. 6. Generation of new *ABO* alleles through gene conversion. In this proposed scheme, the donor alleles are shown in thick lines on the left, the acceptor alleles in thin lines in the middle and the products of gene conversion on the right. The vertical lines represent the SNP sites in the corresponding alleles. The '+' above the alleles *O211* and *O204* represents additional new base substitutions in these alleles.

grounds. For instance, the base substitution 646T>A is found on the *A101* background to give *Ax01*, on the *O101* background to produce *O113* and on the *O111* background to form *O112* (Figs 1, 3). One could explain this by proposing that the substitution 646T>A occurred repeatedly on different haplotypes to produce new alleles. That this substitution is a transversion makes this hypothesis less plausible. Similarly, the two *O201*-defining substitutions 646T>A and 681G>A are found on the *A101* background to give *Ax04*, on the *A102* background to form *Ael02* and on the *O101* background to produce *O114* (Figs 1, 3). One could also try to explain their occurrence by proposing that these two mutations occurred in succession and repeatedly on different haplotypes to form new alleles. This hypothesis is even less probable than the first one because this involves two mutations occurring in succession. In both scenarios, double recombination occurring within such a short DNA segment also seems very unlikely.

In *B107*, the base substitution 526C>G characteristic of *B101* is not found and instead the sequence at this position is the same as that in *A101* (Fig. 2). The applies to the nt 657 of *B103* and the nt 657 and 702 of *B(A)01*. One explanation is that these alleles are the intermediate forms in the process of forming a typical *B101*. But this means there have to be many different intermediate forms and several sequential orders of base substitutions. Thus, this explanation is not so plausible. Another explanation is that backward mutation occurred at these positions in the *B101* allele. For each backward mutation, there are three possible outcomes. Take the nt 526 as an example. At this position, the base is G in *B101*. Substitution can result in A, C or T and only a substitution of C for G can give *B107*. The same argument applies to *B103*, *B(A)01* and other similar *B* alleles. But this does not explain why such backward mutations occurring at the *B101*-defining positions always produce sequences identical to that of *A101*.

The occurrence of this group of alleles can elegantly be explained using a unifying hypothesis if one invokes gene conversion to account for their genesis. Gene conversion refers to a non-reciprocal transfer of genetic information in which the sequence of one DNA strand (acceptor sequence) is altered so as to become identical to the sequence of another DNA strand (donor sequence). Note that the donor sequence remains unchanged. The haplotypes of the genetic markers flanking either side of the converted region in the acceptor sequence also remain unaltered after gene conversion. Figure 6 shows the *ABO* alleles explicable by gene conversion. The donor alleles are common alleles and include *O201*, *O101*, *A101*, *A102* and *B101*. Acceptor alleles are mostly common alleles while some are rare alleles (e.g. *A205* and *O111*). Note that over half of the proposed gene conversion events occurred around nt 646–681. It is thus probable that this region is a hotspot of gene conversion, as has also been proposed by Ogasawara *et al.* (2001) on the basis of four putatively gene-converted alleles. Gene conversion is in fact very important in generating the genetic diversity of many loci. For example, it can create new alleles in the glycophorin genes encoding the MNSs blood group antigens, the RH blood group genes and the major histocompatibility loci (Blumenfeld & Huang, 1995; Carritt *et al.* 1997; Högstrand & Böhme, 1999).

O119 could arise through two gene conversion events involving *O201* as the donor allele and *O101* as the acceptor allele (Fig. 6). But it could more likely be generated by recombination between *O101* and *O208*, as has been proposed by Roubinet *et al.* (2001). *O208* itself could be produced by a single gene conversion event involving *O201* as the acceptor allele and *O101* (or *A101* or *A102*) as the donor allele (Fig. 6).

METHODS FOR GENOTYPING THE *ABO* LOCUS

Molecular cloning of the *ABO* gene and elucidation of the molecular basis of its major alleles has allowed the direct determination of the *ABO* genotypes without family studies. Coupled with DNA sequencing, this allows the examination and analysis of the *ABO* locus at the

DNA level in great detail. A fruitful outcome of all these studies is the identification of a large number of new *ABO* alleles, as has been discussed above. There are now many methods available for *ABO* genotyping.

The early methods distinguished the three major alleles (*A*, *B* and *O*) by restriction analysis of SNPs at nt 261 plus at least one other SNP to differentiate between *B* and non-*B* alleles (e.g. nt 526, 703 or 796), with or without multiplexing the PCRs (Yamamoto & Hakomori, 1990*a*; Lee & Chang, 1992; O'Keefe & Dobrovic, 1993; Stroncek *et al*. 1995). If two or more of these *B* versus non-*B* SNPs were assayed, the *O* allele could further be sub-divided into *O101* and *O303* (Yamamoto & Hakomori, 1990*a*; Grunnet *et al*. 1994; Mifsud *et al*. 1996). Restriction analysis of SNPs at nt 261 and 297 within a single PCR fragment allowed the distinction of four alleles (*A*, *B*, *O101* and *O201*) (Akane *et al*. 1996). This approach culminated in the development of a single-tube single-lane genotyping method claimed to discriminate five alleles (*A101*, *A201*, *B101*, *O101* and *O303*): simultaneous double digestion of duplex PCR products assayed SNPs at nt 261, 467, 703 and 1096 (Olsson & Chester, 1995). However, this particular method solely relied on the SNP at nt 467 to differentiate between *A101* and *A201*, and thus could not distinguish between *A102* and *A201* (Fig. 1). This fault would apply to any method that types the *A201* allele exclusively on the basis of the SNP at nt 467 (O'Keefe & Dobrovic, 1996). Note that *A102* is common in Orientals and less frequent in Caucasians, while *A201* is frequent in Caucasians but rare in Orientals (see above).

Another approach employs allele specific amplification. The three major alleles (*A*, *B* and *O*) were discriminated by allele specific amplification using primers of different lengths and targeting three SNP sites, namely, nt 261, 526 and 703 (Ugozzoli & Wallace, 1992). One report described an elegant use of this approach to distinguish six alleles (*A101*, *A201*, *B101*, *O101*, *O201* and *O303*) by mixing 10 primers in a single reaction to produce PCR products of different lengths and targeting five SNP sites at nt 261, 297, 796, 802 and 1059 (Watanabe *et al*. 1997). Another protocol identified five alleles (*A101*, *A201*, *B101*, *O101* and *O201*) by means of eight separate allele specific PCRs, two for each of the four assayed SNP sites at nt 261, 802, 803 and 1059 (Gassner *et al*. 1996). This was further improved by multiplexing the reactions so that only two separate multiplex allele specific PCRs were required (Pearson & Hessner, 1998). Still another approach was the combined use of allele specific amplification and restriction analysis: five alleles (*A101*, *A102*, *B101*, *O101* and *O201*) were thus differentiated on the basis of seven SNPs (nt 261, 467, 526, 646, 703, 796 and 803) in three separate PCR products (Fukomori *et al*. 1996; Kang *et al*. 1997). The *O303* allele could also be distinguished by assaying one additional SNP at nt 802 in another PCR product (Nishimukai *et al*. 1996).

The fourth approach utilizes mutation screening methods that detect both known and unknown SNPs in the assayed PCR products, e.g. denaturing gradient gel electrophoresis (Johnson & Hopkinson, 1992; Yip *et al*. 1995, 1996) and single strand conformation polymorphism (SSCP) analysis (Akane *et al*. 1996; Ogasawara *et al*. 1996*a*; Tsai *et al*. 2000; Yip, 2000). Four alleles (*A*, *B*, *O101* and *O201*) could be discriminated by SSCP analysis of a single PCR product amplified from exon 6 (Akane *et al*. 1996). The same four alleles could also be determined by amplifying and analysing two separate PCR products under the same conditions (Tsai *et al*. 2000). The analysis was based on simultaneous SSCP and heteroduplex analysis using commercial precast polyacrylamide gels with the detection of heterozyotes being enhanced by the presence of heteroduplex bands. It is very exciting to note that 13 different alleles (common and rare) could be identified by SSCP analysis of four separate PCR products amplified from exons 6 and 7 (Ogasawara *et al*. 1996*a*). This approach was further improved by multiplexing three PCRs in a single tube and analysing the products with SSCP in a single lane (Yip, 2000). This multiplex PCR-SSCP protocol assayed simultaneously the well-established base changes at

nine positions (nt 261, 297, 467, 526, 646, 657, 681, 1059 and 1096) so that seven common alleles (*A101*, *A102*, *A201*, *B101*, *O101*, *O201* and *O303*) could be distinguished using a single-tube and single-lane format. The redundancy in the known SNPs being assayed increased the accuracy and reliability of the genotyping method. This technique is the simplest, quickest and most informative method reported to date, and also readily identifies new alleles because of the capability of SSCP analysis to detect unknown SNPs in the fragments.

Another completely different approach is based on measuring the consumption of labelled primers in PCR (Watanabe *et al*. 2001). The consumed allele-specific primer analysis involves the use of labelled allele-specific primers differing in length and targeting the SNPs at nt 261 and 803. The amount of the primers remaining after PCR are measured to allow the determination of the three major alleles (*A*, *B* and *O*).

CONCLUDING REMARKS

The large variety of methods available for genotyping the *ABO* locus has allowed the identification of over 70 alleles in the last 10 years. Though many of these alleles are rare and may only be found in the populations under study, careful analysis of these alleles reveals that both new point mutations and genetic recombination (crossovers and, very probably, gene conversion) are equally important in generating the genetic diversity in the *ABO* locus. It is expected that many more new alleles will be reported in the future. In particular, if sequence analysis is extended into coding region other than exons 6 and 7, and non-coding regions like the promoter region and the upstream minisatellite repeats, sequence variation may be identified in these regions. Such sequence variation may explain some of the interesting alleles reported in the literature. These interesting alleles may be typed as certain alleles on the basis of their sequences in exons 6 and 7, but the phenotypes are not the expected ones. For example, some apparent A^2 alleles may produce phenotype A_3 or even A_x. Reduction of glycosytransferase enzymatic activity can be due to abnormal localization of the enzyme in the Golgi apparatus as a result of amino acid substitution in the transmembrane domain. This is exemplified by the deficiency of *FUT3*-encoded fucosyltransferase due to such an amino acid substitution (Mollicone *et al*. 1994).

The author's work on the ABO blood group genetics was funded by grants (G-S062 and G-T083) from the Department of Nursing and Health Sciences, The Hong Kong Polytechnic University. The author also wishes to thank Professor Sue Povey (University College London) for encouraging him to write this review and for her continual support.

REFERENCES

Amado, M., Bennett, E. P., Carneiro, F. & Clausen, H. (2000). Characterization of the histo-blood group O^2 gene and its protein product. *Vox Sang.* **79**, 219–226.

Akane, A., Yoshimura, S., Yoshida, M., Okii, Y., Watabiki, T., Matsubara, K. & Kimura, K. (1996). ABO genotyping following a single PCR amplification. *J. Forensic Sci.* **41**, 272–274.

Barjas-Castro, M. L., Carvalho, M. H., Locatellei, M. F., Bordin, S. & Saad, S. T. O. (2000). Molecular heterogeneity of the A_3 subgroup. *Clin. Lab. Haem.* **22**, 73–78.

Beckers, T., van Loghem, J. J. & Dunsford, I. (1955). A second example of the weak antigen A_4 in the offspring of group O parents. *Vox Sang.* **5**, 145–147.

Bennett, E. P., Steffensen, R., Clausen, H., Weghuis, D. O. & van Kessel, A. G. (1995). Genomic cloning of the human histo-blood group ABO locus. *Biochem. Biophys. Res. Commun.* **206**, 318–325.

Blumenfeld, O. O. & Huang, C. H. (1995). Molecular genetics of the glycophorin gene family, the antigens for MNSs blood groups: multiple gene rearrangements and modulation of splice site usage result in extensive diversification. *Hum. Mutat.* **6**, 199–209.

Carritt, B., Kemp, T. J. & Poulter, M. (1997). Evolution of the human RH (rhesus) blood group genes: a 50 year old prediction (partially) fulfilled. *Hum. Mol. Genet.* **6**, 843–850.

Clausen, H., White, T., Takio, K., Titani, K., Stroud, M., Holmes, E., Karkov, J., Thim, L. & Hakomori, S. (1990). Isolation to homogeneity and partial characterization of a histo-blood group A defined Fuc alpha 1-2Gal alpha 1-3-N-acetylgalactosaminyltransferase from human lung tissue. *J. Biol. Chem.* **265**, 1139–1145.

Colley, K. J. (1997). Golgi localization of glycosyltransferases: more questions than answers. *Glycobiology* **17**, 1–13.

Cooper, D. N. & Krawczak, M. (1993). *Human Gene Mutation*. Oxford: Bios Scientific.

Daniels, G. (1995). *Human Blood Groups*. Oxford: Blackwell Science.

Franco, R. F., Simões, B. P. & Zago, M. A. (1995). Relative frequencies of the two O alleles of the

histo-blood ABH system in different racial groups. *Vox Sang.* **69**, 50–52.

Ferguson-Smith, M. A., Aitken, D. A., Turleau, C. & de Grouchy, J. (1976). Localisation of the human ABO: Np-1: AK-1 linkage group by regional assignment of AK-1 to 9q34. *Hum. Genet.* **34**, 35–43.

Fukumori, Y., Ohnoki, S., Shibata, H. & Nishimukai, H. (1996*a*). Suballeles of the ABO blood group system in a Japanese population. *Hum. Hered.* **46**, 85–91.

Fukumori, Y., Ohnoki, S., Yoshimura, K., Nagao, N., Shibata, H., Tomita, T., Okubo, Y. & Yamaguchi, H. (1996*b*). Rapid detection of the cis-AB allele consisting of a chimera of normal A and B alleles by PCR-RFLPs. *Transfus. Med.* **6**, 337–344.

Gassner, G., Schmarda, A., Nussbaumer, W. & Schönitzer, D. (1996). ABO glycosyltransferase genotyping by polymerase chain reaction using sequence-specific primers. *Blood* **88**, 1852–1856.

Grunnet, N., Steffensen, R., Bennett, E. P. & Clausen, H. (1994). Evaluation of histo-blood group ABO genotyping in a Danish population: frequency of a novel O allele defined as O^2. *Vox Sang.* **67**, 210–215.

Högstrand, K. & Böhme, J. (1999). Gene conversion can create new MHC alleles. *Immunol. Rev.* **167**, 305–317.

Hosoi, E. & Yoshikawa, K. (1993). Genetic analysis of the genotypes of ABO and cis-AB blood groups. *Rinsyo Byori* **41**, 1133–1140.

The International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933.

Irshaid, N. M., Chester, M. A. & Olsson, M. L. (1999). Allele-related variation in minisatellite repeats involved in the transcription of the blood group ABO gene. *Transfus. Med.* **9**, 219–226.

Kang, S. H., Fukumori, Y., Ohnoki, S., Shibata, H., Han, K. S., Nishimukai, H. & Okubo, Y. (1997). Distribution of ABO genotypes and allele frequencies in a Korean population. *Jpn. J. Hum. Genet.* **42**, 331–335.

Kennedy, G. C., German, M. S. & Rutter, W. J. (1995). The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat. Genet.* **9**, 293–298.

Kominato, F., Hata, Y., Takizawa, H., Tsuchiya, T., Tsukada, J. & Yamamoto, F. (1999). Expression of human histo-blood group ABO genes is dependent upon DNA methylation of the promoter region. *J. Biol. Chem.* **274**, 37240–37250.

Kominato, Y., Tsuchiya, T., Hata, N., Takizawa, H. & Yamamoto, F. (1997). Transcription of human ABO histo-blood group genes is dependent upon binding of transcription factor CBF/NF-Y to minisatellite sequence. *J. Biol. Chem.* **272**, 25890–25898.

Lee, J. C. I. & Chang, J. G. (1992). ABO genotyping by polymerase chain reaction. *J. Forensic Sci.* **37**, 1269–1275.

van Loghem, J. J. & van der Hart, M. (1954). The weak antigen A4 occurring in the offspring of group O parents. *Vox Sang.* **4**, 69–75.

Maity, S. N. & de Crombrugghe, B. (1998). Role of the CCAAT-binding protein CBF/NF-Y in transcription. *TIBS* **23**, 174–178.

Mifsud, N. A., Haddad, A. P., Condon, J. A. & Sparrow, R. L. (1996). ABO genotyping by polymerase chain reaction-restriction fragment length polymorphism. *Immunohematology* **12**, 143–148.

Mifsud, N. A., Watt, J. M., Condon, J. A., Haddad, A. P. & Sparrow, R. L. (2000). A novel cis-AB variant allele arising from a nucleotide substitution A796C in the transferase gene. *Transfusion* **40**, 1276–1277.

Mollicone, R., Reguigne, I., Kelly, R. J., Fletcher, A., Watt, J., Chatfield, S., Aziz, A., Cameron, H. S., Weston, B. W. & Lowe, J. B. (1994). Molecular basis for Lewis alpha(1,3/1,4)-fucosyltransferase gene deficiency (FUT3) found in Lewis-negative Indonesian pedigrees. *J. Biol. Chem.* **269**, 20987–20994.

Nishimukai, H., Fukumori, Y., Okiura, T., Yuasa, I., Shinomiya, T., Ohnoki, S., Shibata, H. & Vogt, U. (1996). Genotyping of the ABO blood group system; analysis of nucleotide position 802 by PCR-RFLP and the distribution of ABO genotypes in a German population. *Int. J. Legal Med.* **109**, 90–93.

Ogasawara, K., Bannai, M., Saitou, N., Yabe, R., Nakata, K., Takenaka, M., Fujisawa, K., Uchikawa, M., Ishikawa, Y., Juji, T. & Tokunaga, K. (1996*a*). Extensive polymorphism of ABO blood group gene: three major lineages of the alleles for the common ABO phenotypes. *Hum. Genet.* **97**, 777–783.

Ogasawara, K., Yabe, R., Uchikawa, M., Saitou, N., Bannai, M., Nakata, K., Takenaka, M., Fujisawa, K., Ishikawa, Y., Juji, T. & Tokunaga, K. (1996*b*). Molecular genetic analysis of variant phenotypes of the ABO blood group system. *Blood* **88**, 2732–2737.

Ogasawara, K., Yabe, R., Uchikawa, M., Bannai, M., Nakata, K., Takenaka, M., Takahashi, Y., Juji, T. & Tokunaga, K. (1998). Different alleles cause an imbalance in A_2 and A_2B phenotypes of the ABO blood group. *Vox Sang.* **74**, 242–247.

Ogasawara, K., Yabe, R., Uchikawa, M., Nakata, K., Watanabe, J., Takahashi, Y. & Tokunaga, K. (2001). Recombination and gene conversion-like events may contribute to ABO gene diversity causing various phenotypes. *Immunogenetics* **53**, 190–199.

O'Keefe, D. S. & Dobrovic, A. (1993). A rapid and reliable PCR method for genotyping the ABO blood group. *Hum. Mutat.* **2**, 67–70.

O'Keefe, D. S. & Dobrovic, A. (1996). A rapid and reliable PCR method for genotyping the ABO blood group. II. A^2 and O^2 alleles. *Hum. Mutat.* **8**, 358–361.

Olsson, M. L. & Chester, M. A. (1995). A rapid and simple ABO genotype screening method using a novel B/O^2 versus A/O^2 discriminating nucleotide substitution at the ABO locus. *Vox Sang.* **69**, 242–247.

Olsson, M. L. & Chester, M. A. (1996*a*). Frequent occurrence of a variant O^1 gene at the blood group ABO locus. *Vox Sang.* **70**, 26–30.

Olsson, M. L. & Chester, M. A. (1996*b*). Polymorphisms at the ABO locus in subgroup A individuals. *Transfusion* **36**, 309–313.

Olsson, M. L. & Chester, M. A. (1996*c*). Evidence for a new types of *O* allele at the *ABO* locus, due to a combination of the A^2 nucleotide deletion and the A^{e1} nucleotide insertion. *Vox Sang.* **71**, 113–117.

Olsson, M. L. & Chester, M. A. (1998). Heterogeneity of the blood group A^x allele: genetic recombination of common alleles can result in the Ax phenotype. *Transfus. Med.* **8**, 231–238.

Olsson, M. L., Guerreiro, J. F., Zago, M. A. & Chester, M. A. (1997). Molecular analysis of the O alleles at the blood group ABO locus in populations of different ethnic origin reveals novel crossing-over events and point mutations. *Biochem. Biophys. Res. Commun.* **234**, 779–782.

Olsson, M. L., Santos, S. E. B., Guerreiro, J. F., Zago, M. A. & Chester, M. A. (1998). Heterogeneity of the O alleles at the blood group ABO locus in Amerindians. *Vox Sang.* **74**, 46–50.

Olsson, M. L., Thuresson, B. & Chester, M. A. (1995). An A^{EL} allele-specific nucleotide insertion at the blood group ABO locus and its detection using a sequence-specific polymerase chain reaction. *Biochem. Biophys. Res. Commun.* **216**, 642–647.

Pearson, S. L. & Hessner, M. J. (1998). $A^{1,2}BO^{1,2}$ genotyping by multiplexed allele-specific PCR. *Brit. J. Haematol.* **100**, 229–234.

Roubinet, F., Kermarrec, N., Despiau, S., Apoil, P. A., Dugoujon, J. M. & Blancher, A. (2001). Molecular polymorphism of O alleles in five populations of different ethnic origins. *Immunogenetics* **53**, 95–104.

Seto, N. O. L., Palcic, M. M., Compston, C. A., Li, H., Bundle, D. R. & Narang, S. A. (1997). Sequential interchange of four amino acids from blood group B to blood group A glycosyltransferase boosts catalytic activity and progressively modifies substrate recognition in human recombinant enzymes. *J. Biol. Chem.* **272**, 14133–14138.

Smith, G. R. (1994). Hotspots of homologous recombination. *Experentia* **50**, 234–241.

Stroncek, D. F., Konz, R., Clay, M. E., Houchins, J. P. & McCullough, J. (1995). Determination of ABO glycosyltransferase genotypes by use polymerase chain reaction and restriction enzymes. *Transfusion* **35**, 231–240.

Suzuki, K., Iwata, M., Tsuji, H., Takagi, T., Tamura, A., Ishimoto, G., Ito, S., Matsui, K. & Miyazaki, T. (1997). A de novo recombination in the ABO blood group gene and evidence for the occurrence of recombination products. *Hum. Genet.* **99**, 454–461.

Tsai, L. C., Kao, L. G., Chang, J. G., Lee, H. H., Linacre, A. & Lee, J. C. I. (2000). Rapid identification of the ABO genotypes by their single-stand conformation polymorphism. *Electrophoresis* **21**, 537–540.

Ugozzoli, L. & Wallace, B. (1992). Application of an allele-specific polymerase chain reaction to the direct determination of ABO blood group genotypes. *Genomics* **12**, 670–674.

Watanabe, G., Umetsu, K., Yuasa, I. & Suzuki, T. (1997). Amplified product length polymorphism (APLP): a novel strategy for genotyping the ABO blood group. *Hum. Genet.* **99**, 34–37.

Watanabe, G., Umetsu, K., Yuasa, I., Sato, M., Sakabe, M., Naito, E., Yamanouchi, H. & Suzuki, T. (2001). A novel technique for detecting single nucleotide polymorphisms by analyzing consumed allele-specific primers. *Electrophoresis* **22**, 418–420.

Watkins, W. M. (1980). Biochemistry and genetics of the ABO, Lewis, and P blood group system. In *Advances in Human Genetics* (eds. Harris, H. & Hirschhorn, K.), Vol. 10, pp. 1–136. New York: Plenum Press.

Yamamoto, F. (2000). Molecular genetics of ABO. *Vox Sang.* **78** (suppl 2), 91–103.

Yamamoto, F. & Hakomori, S. (1990). Sugar-nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. *J. Biol. Chem.* **265**, 19257–19262.

Yamamoto, F., Clausen, H., White, T., Marken, J. & Hakomori, S. (1990*a*). Molecular genetic basis of the histo-blood group system. *Nature* **345**, 229–233.

Yamamoto, F., Marken, J., Tsuji, T., White, T., Clausen, H. & Hakomori, S. (1990*b*). Cloning and characterization of DNA complementary to human UDP-GalNAc: Fucα1$\rightarrow$2Galα1$\rightarrow$3GalNAc transferase (histo-blood group A transferase) mRNA. *J. Biol. Chem.* **265**, 1146–1151.

Yamamoto, F. & McNeill, P. D. (1996). Amino acid residue at codon 268 determines both activity and nucleotide-sugar donor substrate specificity of human histo-blood group A and B transferases. *J. Biol. Chem.* **271**, 10515–10520.

Yamamoto, F., McNeill, P. D. & Hakomori, S. (1992). 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. *Biochem. Biophys. Res. Commun.* **187**, 366–374.

Yamamoto, F., McNeill, P. D. & Hakomori, S. (1995). Genomic organization of human histo-blood group ABO genes. *Glycobiology* **5**, 51–58.

Yamamoto, F., McNeill, P. D., Kominato, Y., Yamamoto, M., Hakomori, S., Ishimotot, S., Nishida, S., Shima, M. & Fujimura, Y. (1993*a*). Molecular genetic analysis of the ABO blood group system: 2. Cis-AB alleles. *Vox Sang.* **64**, 120–123.

Yamamoto, F., McNeill, P. D., Yamamoto, M., Hakomori, S., Bromilow, I. M. & Duguid, J. K. M. (1993*b*). Molecular genetic analysis of the ABO blood group system: 4. Another type of O allele. *Vox Sang.* **64**, 175–178.

Yamamoto, F., McNeill, P. D., Yamamoto, M., Hakomori, S. & Harris, T. (1993*c*). Molecular genetic analysis of the ABO blood group system: 3. A^x and $B^{(A)}$ alleles. *Vox Sang.* **64**, 171–174.

Yamamoto, F., McNeill, P. D., Yamamoto, M., Hakomori, S., Harris, T., Judd, W. J. & Davenport, R. D. (1993*d*). Molecular genetic analysis of the ABO blood group system: 1. Weak subgroups: A^3 and B^3 alleles. *Vox Sang.* **64**, 116–119.

Yip, S. P. (2000). Single-tube multiplex PCR-SSCP analysis distinguishes 7 common ABO alleles and readily identifies new alleles. *Blood* **95**, 1487–1492.

Yip, S. P., Choy, W. L., Chan, C. W. & Choi, C. H. (1996). The absence of a B allele in acquired B blood group phenotype confirmed by a DNA based genotyping method. *J. Clin. Pathol.* **49**, 180–181.

Yip, S. P., Yow, C. M. N. & Lewis, W. H. P. (1995). DNA polymorphism at the ABO locus in the Chinese population of Hong Hong. *Hum. Hered.* **45**, 266–271.

Yu, L. C., Lee, H. L., Chan, Y. S. & Lin, M. (1999). The molecular basis for the B(A) allele: an amino acid alteration in the human histoblood group B α-(1,3)-galactosyltransferase increases its intrinsic α-(1,3)-N-acetylgalactosaminyltransferase activity. *Biochem. Biophys. Res. Commun.* **262**, 487–493.

Yu, L. C., Chang, C. Y., Twu, Y. C. & Lin, M. (2000). Human histo-blood group ABO glycosyltransferase genes: different enhancer structures with different transcriptional activities. *Biochem. Biophys. Res. Commun.* **273**, 459–466.