BIOCHEMICAL LETTERS JOURNAL

The two-gene model of the RH blood-group locus

Our knowledge of the biochemistry and molecular genetics of the RH blood-group system has made considerable progress recently (for reviews see [1,2]). This was the result of numerous studies on the chemistry of Rh proteins, which lead to the successful cloning by two independent groups of the first Rh cDNA clones [3,4], later recognized as complementary copies of a transcript from the *RHCE* gene [5–7]. The identification of the cDNA encoding the RhD protein was accomplished more recently [6], following a careful characterization of several cDNAs from a human bone-marrow library, together with simultaneous cloning and sequencing of genomic DNA libraries from RhD-positive and RhD-negative individuals. Accordingly, the RhXIII cDNA clone with 41 nucleotide substitutions was isolated, and its sequence matched perfectly with the nucleotide sequence of genomic fragments isolated from RhD-positive, but not RhD-negative, individuals [6]. This comparison was necessary, since the deduced RhD protein sequence was different (down to position 42) from that found by protein sequencing the first 54 amino acids of a purified RhD protein [4]. Although reliable sequencing of a very long stretch of peptide is difficult, there is at present no clear explanation for the discrepancy between results obtained by cloning [6] and protein sequencing of RhD [4], but our findings and sequence were confirmed later by another group [8], which identified a cDNA clone (Rh13), identical with RhXIII, that was the product of a gene present in RhD-positive, but not RhD-negative, individuals. The polypeptides encoded by the RhCE and RhD cDNA clones exhibit 96% similarity and differ by 35 amino acid substitutions. They are thought to derive from two genes that evolved by duplication.

In parallel with these investigations, Southern-blot analysis carried out with genomic DNA from Caucasian individuals serotyped as RhD-positive and RhD-negative (whatever the C/c and E/e phenotypes), digested with several restriction enzymes and exon-specific probes for the *RHCE* gene, clearly indicated that the RH locus of humans is characterized by a polymorphism in the number of structural genes [9]. The RhD-positive chromosome carries two closely linked genes, *RHD* and *RHCE*, at position 1p34–p36 [9] that define an haplotype which is inherited *en bloc* from one generation to another, as expected from familial studies [10]. Conversely, the RhD-negative chromosome carries a single gene, *RHCE*, as none of the RhD specific sequences could be detected with full-length or exon-specific probes in RhD-negative individuals [9]. Thus the *RHD* gene should encode the RhD protein and the *RHCE* gene should encode the C/c and E/e proteins, most likely following alternative splicing of a primary transcript [11]. As discussed below, this model has been largely confirmed by further studies of Rh transcripts from individuals of different Rh phenotypes.

Two recent papers published in this Journal [12,13] raise a confusion and suggest that the RH locus is composed of three genes and that the cDNA clones RhXIII [6] and Rh13 [8] are not the products of the *RHD* gene. Unfortunately, the conclusions of

these authors [12,13] are not supported by their data. (i) A new genomic polymorphism associated with the RhC/c phenotypes unrelated to the RhD-positive and RhD-negative phenotypes was detected by *RsaI* restriction pattern and, surprisingly, this was argued to conclude that ‘this new polymorphism may support the three-gene theory rather than the two-gene theory’ [12]. Obviously, the two-gene model is fully compatible with C/c polymorphisms, as demonstrated by others [14,15]. Moreover, we have used the genetic polymorphism associated with C/c [16] to settle a prenatal diagnosis in a fetus at risk of haemolytic disease of the newborn [17]. (ii) The Japanese group claimed that a cDNA clone (RhPII) which is almost identical with RhXIII and Rh13 could be amplified by PCR from reticulocyte RNA and genomic DNA of RhD-positive and RhD-negative individuals, suggesting that RhPII is not the product of the *RHD* gene [13,18]. Although this assumption may appear correct, it is in contradiction with many other arguments (see below) and it may simply indicate the detection of a new allele in a population where the Rh-negative chromosome is rare. The observed dissociation between serotype and genotype would also be explained under the two-gene model, as discussed below.

The two-gene model is further supported by the routine determination of fetal RhD genotype early in pregnancy to eliminate the risks to an RhD-negative fetus of fetal blood sampling or serial amniocentesis [19]. This typing has also been used for single-cell analysis of the RhD type for use in pre-implantation diagnosis in the prevention of severe haemolytic disease of the newborn [20]. More than 500 RhD genotypings of amniotic cells have been carried out by us and others [19–22], with only a very few discordances between serotypes and genotypes. In most instances, a dissociation was observed with weak D variants such as D^u (low grade) or D^{vi}, which are known to be very weakly, or not at all, reactive with most anti-D antibodies [22,23]. In addition, it is expected that some Rh_{null} samples that carry normal *RHD* and *RHCE* genes [14,24], but do not express any of the Rh antigens, would be genotyped as RhD-positive. It is possible that other rare D variants may carry partial deletions that preclude RhD genotyping. For instance, a frameshift or a nonsense mutation in the *RHD* gene may ‘knock out’ this gene, which would not produce any RhD protein but would be genotyped as RhD-positive; this would represent an unusual form of ‘Rh-negative’ phenotype. These findings and considerations indicate that, under very rare circumstances, samples serotyped as D-positive or D-negative may be genotyped as D-negative or D-positive, respectively. During the editorial procedure, these expectations have been confirmed, since fragments of the *RHD* gene and a non-functional *RHD* gene have been identified in some Rh-negative individuals from Australia [25] and from Blacks [26]. Therefore the Rh-negative phenotype is heterogeneous, and a careful analysis of the RH locus according to the ethnic origin of the samples should be carried out. It is unlikely that these findings will limit the use of the RhD typing by PCR in fetuses, because the variants discussed above are rare, and in most occasions they will give a ‘false positive’ answer that will have no consequence for mother immunization. However, a ‘false negative’ answer is potentially possible with D variants

Figure 1 Structure of the human blood-group Rh locus

Schematic representation showing the two structural genes (*RHD* and *RHCE*) identified in the genome of RhD-positive individuals (D⁺). The RhD-negative individuals are heterogeneous: (1) in most donors of Caucasian origin, only the *RHCE* gene is present and functional; (2) in some Australians and Blacks, a fragment of the *RHD* gene or a non-functional *RHD* gene is present.

arising from gene deletion or gene conversion between the homologous *RHD* and *RHCE* genes [23,26a]. Although these variants are rare, they are potentially immunogenic. The use of two pairs of primers probing different regions of the *RHD* gene [21,22] would largely decrease (or eliminate?) the number of these situations.

Two other results strongly argue in favour of the two-gene model. (1) Genomic analysis of D-category phenotypes that produce anti-D antibodies [27] indicates that the loss of D epitopes from RhD proteins was associated with rearrangements of the *RHD* gene. Indeed, segmental DNA replacements (gene conversions) of D-specific sequences by CE-specific sequences have been identified in all instances [23,26a]. As a result, Rh 'D-CE-D' hybrid proteins lacking one or several of the nine epitopes that composed the D mosaic structure are generated. These data demonstrate that alterations of the *RHD* gene (defined by the RhXIII cDNA clone and in Figure 1) are strictly correlated with alterations of the RhD antigens carried by circulating erythrocytes. (2) The finding that a same transcript isolated from human reticulocytes carries the genetic mutations associated with C/c and E/e polymorphisms [16] represents to us the definitive proof that the C/c and E/e changes are encoded by the same gene (*RHCE*), although the C/c and E/e proteins might be carried by different polypeptides [28], presumably generated by alternative splicing of a single pre-messenger RNA [11].

In conclusion, both protein and molecular-genetic analysis are fully compatible with a two-gene model of the RH locus (Figure 1). Conversely, there is no experimental evidence for the presence of a third gene at the RH locus. The two-gene model may therefore serve as a model for future analysis of the structure and

regulation of the *RHD* and *RHCE* genes in normal and pathological situations.

Jean-Pierre CARTRON, Caroline LE VAN KIM, Baya CHERIF-ZAHAR, Isabelle MOURO, Christelle ROUILLAC and Yves COLIN

Unité INSERM U76, Institut National de la Transfusion Sanguine, 6 rue Alexandre Cabanel, 75015 Paris, France

- Cartron, J. P. and Agre, P. (1993) *Semin. Hematol.* **30**, 193–208
- Anstee, D. J. and Tanner, M. J. A. (1993) *Bailliere's Clin. Haematol.* **6**, 401–422
- Chérif-Zahar, B., Bloy, C., Le Van Kim, C., Blanchard, D., Bailly, P., Hermand, P., Salmon, C., Cartron, J. P. and Colin, Y. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6243–6247
- Avent, N. D., Ridgwell, K., Tanner, M. J. A. and Anstee, D. J. (1990) *Biochem. J.* **271**, 821–825
- Hermand, P., Mouro, I., Huet, M., Bloy, C., Suyama, K., Goldstein, J., Cartron, J. P. and Bailly, P. (1993) *Blood* **82**, 669–676
- Le Van Kim, C., Mouro, I., Chérif-Zahar, B., Raynal, V., Cherrier, C., Cartron, J. P. and Colin, Y. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10925–10929
- Chérif-Zahar, B., Le Van Kim, C., Raynal, V., Cartron, J. P. and Colin, Y. (1994) *Genomics* **19**, 68–74
- Arce, M. A., Thompson, E. S., Wagner, S., Coyne, K. E., Ferdmen, B. A. and Lublin, D. M. (1993) *Blood* **82**, 651–655
- Colin, Y., Chérif-Zahar, B., Le Van Kim, C., Raynal, V., Van Huffel, V. and Cartron, J. P. (1991) *Blood* **78**, 2747–2752
- Race, R. R. and Sanger, R. (1975) *Blood Groups in Man*, 6th edn., pp. 178–260, Blackwell, New York
- Le Van Kim, C., Chérif-Zahar, B., Raynal, V., Mouro, I., Lopez, M., Cartron, J. P. and Colin, Y. (1992) *Blood* **80**, 1074–1078
- Umenishi, F., Kajii, E. and Ikemoto, S. (1994) *Biochem. J.* **299**, 203–206
- Umenishi, F., Kajii, E. and Ikemoto, S. (1994) *Biochem. J.* **299**, 207–211
- Carritt, B., Blunt, T., Avent, N., Daniels, G. and Steers, F. (1993) *Ann. Hum. Genet.* **57**, 273–279
- Hyland, C. A., Wolter, L. C., Liew, Y. W. and Saul, A. (1994) *Blood* **83**, 566–572
- Mouro, I., Colin, Y., Chérif-Zahar, B., Cartron, J. P. and Le Van Kim, C. (1993) *Nature Genetics* **5**, 62–65
- Le Van Kim, C., Mouro, I., Brossard, Y., Chavinié, J., Cartron, J. P. and Colin, Y. (1994) *Br. J. Haematol.* **88**, 193–195
- Kajii, E., Umenishi, F., Iwamoto, S. and Ikemoto, S. (1993) *Hum. Genet.* **91**, 157–162
- Bennett, P., Le Van Kim, C., Colin, Y., Warwick, R., Chérif-Zahar, B., Fisk, N. M. and Cartron, J. P. (1993) *N. Engl. J. Med.* **329**, 607–610
- Van den Veyver, I. B., Chong, S. S., Cota, J., Bennett, P. R., Fisk, N. M., Handyside, A. H., Cartron, J. P., Le Van Kim, C., Colin, Y., Snabes, M. C., Moise, K. J. and Hughes, M. R. (1994) *Am. J. Obstet. Gynecol.*, in the press
- Simsek, S., Bleeker, P. M. M. and von dem Borne, A. E. G. (1994) *N. Engl. J. Med.* **330**, 795–796
- Bennett, P., Warwick, R. and Cartron, J. P. (1994) *N. Engl. J. Med.* **330**, 795–796
- Mouro, I., Le Van Kim, C., Rouillac, C., van Rhenen, D. J., Le Pennec, P. Y., Bailly, P., Cartron, J. P. and Colin, Y. (1994) *Blood* **83**, 1129–1135
- Chérif-Zahar, B., Le Van Kim, C., Raynal, V., D'Ambrosio, A. M., Cartron, J. P. and Colin, Y. (1993) *Blood* **82**, 656–662
- Hyland, C. A., Wolter, L. C. and Saul, A. (1994) *Blood* **84**, 321–324
- Carritt, B., Steers, F. J. and Avent, N. (1994) *Lancet* **344**, 205–206
- Rouillac, C., Colin, Y., Hughes-Jones, N. C., Beolet, M., D'Ambrosio, A. M., Cartron, J. P. and Le Van Kim, C. (1995) *Blood*, in the press
- Tippett, P. (1990) *J. Immunogenet.* **17**, 247–257
- Blanchard, D., Bloy, C., Hermand, P., Cartron, J.-P., Saboori, A., Smith, B. L. and Agre, P. (1988) *Blood* **72**, 1424–1427