Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 2221–2224, March 1994
Genetics

Multiple substitutions in the von Willebrand factor gene that mimic the pseudogene sequence

(gene conversion)

JEROEN C. J. EIKENBOOM*, TOM VINK†, ERNEST BRIËT*, JAN J. SIXMA†, AND PIETER H. REITSMA*

*Hemostasis and Thrombosis Research Center, Department of Hematology, Building 1, C 2-R, University Hospital, P.O. Box 9600, 2300 RC Leiden, The Netherlands; and †Department of Hematology, University Hospital, P.O. Box 85500, 3508GA Utrecht, The Netherlands

Communicated by Kenneth M. Brinkhous, December 1, 1993

ABSTRACT **We have analyzed a type IIB and a type I von Willebrand disease family for the presence of mutations in the region coding for the glycoprotein Ib binding domain of the von Willebrand factor. Since this sequence is also present in the highly homologous von Willebrand factor pseudogene, we have studied genomic DNA as well as cDNA, which was produced from RNA isolated from endothelial cells or platelets. In both families, we have detected multiple consecutive nucleotide substitutions in the 5′ end of exon 28 that result in a sequence identical to the von Willebrand factor pseudogene. These substitutions were also found in cDNA, which proves that they are present in the active gene. The occurrence of multiple adjacent substitutions that exactly reflect a part of the sequence of the von Willebrand factor pseudogene is difficult to reconcile with sequential single mutational events. We therefore hypothesize that each of these multiple substitutions arose from one recombinational event between gene and pseudogene.**

von Willebrand disease (vWD) is characterized by a deficiency or functional defect of the von Willebrand factor (vWF) (1). The human vWF, which is a multimeric protein (2), is encoded by the vWF gene on chromosome 12 (3). A nonfunctional pseudogene that corresponds to the vWF gene sequence from exon 23 to 34, and that has diverged ≈3% from the active gene, is present on chromosome 22 (4).

To date, many molecular defects of the vWF gene causing vWD have been described (reviewed in ref. 5). These defects include point mutations and insertions or deletions of a stretch of nucleotides. The majority of abnormalities are missense mutations that are associated with the functionally abnormal vWD variants types IIA and IIB and the variants with defective factor VIII binding. In severe type III vWD gene deletions, nonsense mutations and defective mRNA expression have been identified as the cause of a quantitative defect of vWF. All these abnormalities in the vWF gene can be understood by assuming single mutational events: a substitution of one nucleotide or the deletion or insertion of one or a stretch of nucleotides.

In this paper, we describe two vWD families in which we have found examples of adjacent, multiple nucleotide substitutions that correspond to the vWF pseudogene sequence. It seems unlikely that these multiple substitutions arose as independent single mutations and by chance alone.

MATERIALS AND METHODS

Patients. Two families were studied, one with type IIB vWD and the other with type I vWD. The type I vWD family was characterized by compound heterozygosity of the symptomatic index cases. This family was recently described (indicated in that paper as family 4) (6). The type IIB family was also previously reported (7).

RNA and Genomic DNA Isolation. Umbilical vein endothelial cells were cultured as described (7). Total RNA was isolated from 20×10^6 endothelial cells using RNazol (Cinna/Biotecx Laboratories, Friendswood, TX) according to the instructions of the manufacturer. Platelet RNA was isolated as described (8). Genomic DNA was prepared from peripheral blood leukocytes according to standard methods.

Amplification of RNA and Genomic DNA. cDNA was synthesized from platelet RNA using primer 374 (Table 1) and 2 μl of this cDNA was used as template in a PCR with primers 3570 and 374A as described (8). Ten micrograms of total RNA, isolated from umbilical vein endothelial cells, was used to synthesize cDNA using primer 67 and a cDNA synthesis kit (Boehringer Mannheim) according to the instructions of the manufacturer. One-tenth of this cDNA was used as a template in a PCR mixture containing 50 mM KCl, 1.5 mM $MgCl_2$, 1 mg of bovine serum albumin per ml, 2.5 units of *Taq* DNA polymerase (Promega), 200 μM dNTPs, and 500 ng of primers 66 and 67 in a 100-μl vol. Thirty-five cycles of 1 min melting at 94°C, 1 min annealing at 55°C, and 3 min extension at 72°C in a Bioexcellence DNA incubator (Biores, Woerden, The Netherlands) were performed.

Genomic DNA was amplified by PCR with the primer pairs 28A-374A, 13-18, and 40-41 listed in Table 1 using 1 μg of genomic DNA as a template. These primers contain mismatches with the vWF pseudogene and therefore amplify gene-specific fragments. To improve gene specificity the fragment produced with primers 28A and 374A was digested with *Bst*XI (Promega) before sequencing.

Cloning and Sequencing. All PCR fragments were directly sequenced by either the dideoxynucleotide chain-termination method (M13 sequencing kit; Boehringer Mannheim) or the method of Bachmann (10) using the primers indicated in Table 1.

The reverse transcriptase PCR product from the endothelial cell-derived RNA was purified on a 1% agarose gel, digested with *Bam*HI and *Eco*RI, and cloned in the vector PTZ18 (Bio-Rad). Sequencing of the clones was performed by the dideoxynucleotide chain-termination method on a double-stranded plasmid DNA using either the forward or reverse M13 primers or internal primers (Table 1).

RESULTS

We have previously described the isolation of human umbilical vein endothelial cells from a newborn with vWD type IIB (7, 11) and the abnormal vWF produced by these endothelial cells. Direct sequence analysis of the amplified coding sequence of the glycoprotein Ib binding domain of vWF using cDNA, derived from RNA isolated from these vWD type IIB

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: vWD, von Willebrand disease; vWF, von Willebrand factor.

Table 1. Oligonucleotide primers

Primer	Sequence (5′ → 3′)	Location	Remark
66	CCAGGATCCGAGGTGGCTGGCCGGCGT	3589–3615, exon 27	*Bam*HI, cDNA PCR
67	AATGAATTCCGATCCTTCCAGGACGAA	4527–4501, exon 28	*Eco*RI, RT, cDNA PCR
374	ATGAACTCCTTGCTCCTGTT	4562–4543, exon 28	RT
3570	TGCGTTGACCCTGAAGACTG	3568–3587, exon 27	cDNA PCR, Seq
374A	TGTCCGATCCTTCCAGGACG	4522–4503, exon 28	cDNA PCR, gDNA PCR
28A	CAGAAGTGTCCACAGGTTCT	24/149–168, intron 27	gDNA PCR, Seq
13	CTCAGAAGTGTCCACAGGTT	24/147–166, intron 27	gDNA PCR
18	CGGTCGATCTTGCTGAAGAT	4121–4102, exon 28	gDNA PCR
40	TGGGAATATGGAAGTCATTG	24/115–134, intron 27	gDNA PCR
41	CCGATCCTTCCAGGACGAAC	4519–4500, exon 28	gDNA PCR
31	CTTTGTGGTCGACATGATGG	3894–3913, exon 28	Seq
32	CCATCATGTCGACCACAAAG	3913–3894, exon 28	Seq
43	AAGGCCTTCGTGCTGAGC	4306–4323, exon 28	Seq
44	CACCAGCGAGGTCTTGAA	4068–4085, exon 28	Seq
1311	CTGAGTTTGAAGTGCTGAAG	3872–3891, exon 28	Seq
4070	AGGTGGCCTCCACCAGCGAG	4058–4077, exon 28	Seq

Underlined nucleotides indicate mutant nucleotides to introduce a specific restriction site; double-underlined nucleotides indicate mismatches with the vWF pseudogene sequence. Nucleotide numbering of coding sequence is according to Bonthron *et al.* (9); intronic sequence is numbered according to Mancuso *et al.* (3). *Bam*HI and *Eco*RI, restriction sites introduced for cloning; RT, primers used in reverse transcriptase cDNA synthesis; cDNA PCR, primers used in PCR of cDNA; gDNA PCR, primers used in PCR of genomic DNA; Seq, primers used in sequence reactions.

endothelial cells, indicated 3 nucleotide substitutions in addition to a common vWF polymorphism (Table 2). Sequence analysis of the cloned reverse transcriptase PCR product confirmed that these substitutions are on a single allele. One of these substitutions, $Arg^{543} \rightarrow Trp$, has previously been identified in a number of vWD type IIB patients (12–17). The other two substitutions, $Val^{466} \rightarrow Gly$ and $Asn^{468} \rightarrow Thr$, are identical with the nucleotide sequence of the vWF pseudogene and would result in vWF missing an N-linked glycosylation site at amino acid 468.

Inadvertent amplification of the pseudogene was excluded, since cDNA was used for amplification with two primers spanning intron 27 and no additional pseudogene-like substitutions were detected. To confirm the results, we have analyzed five members of the vWD type IIB family. Only affected members carry the pseudogene-like substitutions, as shown by the presence of a *Mae* III restriction site. In a nonaffected member and a nonrelated type IIB patient, this restriction site is absent (Fig. 1).

In the type I vWD family, we have detected 3 nucleotide substitutions and one common polymorphism on the same allele in exon 28 (Table 2). One of these three substitutions, G-3789 to A (G3789A), is neutral. The other two changes, C3797T and G3835A, result in the amino acid substitutions $Pro^{503} \rightarrow Leu$ and $Val^{516} \rightarrow Ile$. All three changes are congruent to the vWF pseudogene sequence. The PCR fragment of exon 28 was gene specific since no other pseudogene sequence was seen (Fig. 2). We have also found these pseudogene-like substitutions in cDNA synthesized from platelet-derived RNA (Fig. 3), which confirms that they are located in the active gene. At 16 sites of mismatches between the vWF gene and pseudogene, only the gene sequence was present.

Table 2. Substitutions and polymorphisms

Family type	Substitution		Polymorphism	
	Amino acid	Nucleotide	Amino acid	Nucleotide
IIB	$Val^{466} \rightarrow Gly$	GTT → GGT	$Ala^{618} \rightarrow Thr$	GCC → ACC
	$Asn^{468} \rightarrow Thr$	AAC → ACC		
	$Arg^{543} \rightarrow Trp$	CGG → TGG		
I	$Ser^{500} \rightarrow Ser$	TCG → TCA	$Ala^{618} \rightarrow Thr$	GCC → ACC
	$Pro^{503} \rightarrow Leu$	CCG → CTG		
	$Val^{516} \rightarrow Ile$	GTC → ATC		

All substitutions and polymorphisms indicated for each family are located on the same allele.

DISCUSSION

In the type IIB patient, we have found the $Arg^{543} \rightarrow Trp$ mutation, one of the most frequent type IIB mutations. This mutation has been characterized in functional studies as the cause of a type IIB phenotype (13). Whether the pseudogene-like amino acid substitutions $Val^{466} \rightarrow Gly$ and $Asn^{468} \rightarrow Thr$ on the same allele have any additional deleterious effect on the function of the vWF is not known. As we described before (6), the type I phenotype of the second family under study is caused by compound heterozygosity for a "silent allele" and the allele containing the substitutions $Pro^{503} \rightarrow Leu$ and $Val^{516} \rightarrow Ile$. The $Pro^{503} \rightarrow Leu$ mutation has been described as the cause of enhanced platelet reactivity to ristocetin in type I New York (18).

The majority of mutations described in vWD occur at CpG dinucleotides, which are hot spots for mutations (5). These are all single nucleotide substitutions and only three substitutions have been reported that result in a gene sequence corresponding to the sequence of the vWF pseudogene: a nonsense mutation $Arg^{896} \rightarrow ter$ in a type III patient (19), a candidate missense mutation $Gly^{846} \rightarrow Arg$ in a type IIA patient (20), and a preliminary report on a candidate missense

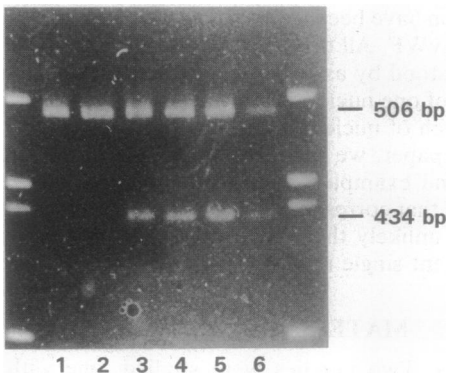

FIG. 1. Genomic DNA amplified with primers 13 and 18, which results in a fragment of 506 bp, was subsequently digested with the restriction enzyme *Mae* III. When the pseudogene-like substitution $Asn^{468} \rightarrow Thr$ is present, the 506-bp fragment will be cleaved into a 434- and a 72-bp fragment. Therefore, the 434-bp fragment should be present only in affected family members (lanes 3–6) of the type IIB vWD family. Lanes: 1, nonrelated type IIB patient; 2, nonaffected family member; 3–6, affected family members.

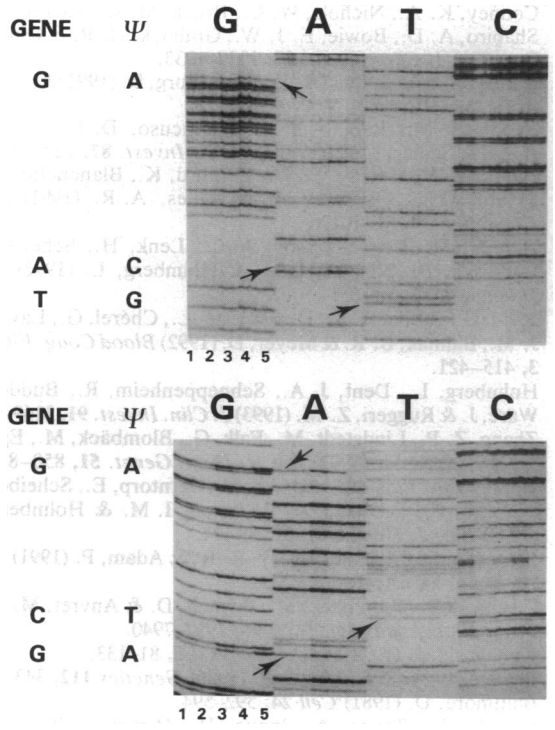

FIG. 2. Direct sequence analysis of genomic DNA from the type I family amplified with primers 28A and 374A and digested with *Bst*XI to eliminate any coamplified pseudogene. (*Upper*) Part of the sequence gel illustrating the gene specificity of the amplification by the absence of the pseudogene sequence. Arrows indicate sites of mismatches between the vWF gene and pseudogene sequence. The expected nucleotides at these three sites are indicated on the left. GENE, vWF gene sequence; ψ, pseudogene sequence. In all lanes, only gene sequence is present (arrows), which indicates gene specificity of the amplified fragment. (*Lower*) Part of the sequence gel showing heterozygosity for three substitutions in the vWF gene that are identical to the expected pseudogene sequence. Arrows indicate substitutions at sites of mismatches between the vWF gene and pseudogene sequence. Expected nucleotides at these three sites of mismatches are indicated on the left. GENE, vWF gene sequence; ψ, pseudogene sequence. Lanes: 1 and 2, compound heterozygous index patients; 3 and 4, family members that have the vWF allele with these pseudogene-like substitutions in common with the patients; 5, family member that carries the second vWF allele of the patients and no pseudogene-like substitutions are present.

mutation Phe^{606} → Ile in an unclassified variant of type I vWD (21). This indicates that, although the vWF gene frequently mutates at hot spots for mutations, it does not preferentially acquire abnormalities that are present in the pseudogene. In two families, we have found multiple adjacent nucleotide substitutions that exactly reflect a small part of the vWF pseudogene sequence. In the same region of the vWF gene, similar multiple substitutions (G3789A and Pro^{503} → Leu) were reported in two families by Holmberg *et al.* (18) and in one family by Zhang *et al.* (22). Although multiple substitutions are rare, they seem to be very frequent in this region of the vWF gene. The substitutions Val^{466} → Gly and Asn^{468} → Thr in the type IIB family and the Val^{516} → Ile substitution in the type I family did not occur at putative hot spots for mutations. It seems unlikely that these multiple adjacent substitutions are the result of sequential single mutational events. We therefore postulate that these base changes result from an alternative mutational mechanism. A candidate concept could be gene conversion.

Gene conversion is characterized by local copying of a nucleotide sequence between homologous DNA sequences (23–25). In our examples, the putatively converted sequences are short: in the type IIB family they are 7–95 bp long and in the type I family they are 47–195 bp long (Fig. 4). Such short gene conversions have been described before in detail for the human fetal globin gene (24). Gene conversion has also been claimed as the mutational mechanism in several other human genes—for example, in steroid 21-hydroxylase deficiency (26), the major histocompatibility genes (27), and the β-globin gene (28). In all these examples, the conversions occur between multiallelic genes or between genes and pseudogenes located on the same chromosome. In contrast, in the case of the vWF gene and pseudogene the conversion is interchromosomal between chromosomes 12 and 22. Interchromosomal gene conversion has recently been postulated for the human genome (29), but it seems to be common in yeast and in *Drosophila* (for review, see ref. 30).

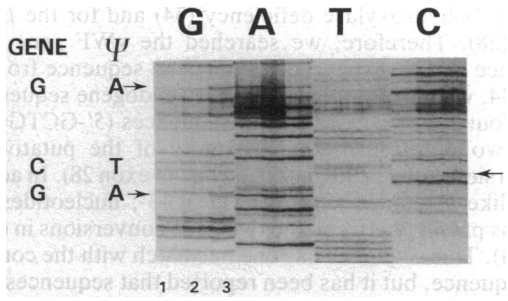

FIG. 3. Part of direct sequence analysis of vWF cDNA of the type I family. cDNA was synthesized through reverse transcription of platelet-derived RNA by using primer 374 and was subsequently amplified by PCR with primers 3570 and 374A, which span intron 27. Arrows indicate substitutions at sites of mismatches between the vWF gene and pseudogene sequence. Expected nucleotides at these three sites of mismatches are indicated on the left. GENE, vWF gene sequence; ψ, pseudogene sequence. Lanes: 1 and 2, compound heterozygous patients; 3, family member sharing the allele with the pseudogene-like substitutions. Patients (lanes 1 and 2) appear to be homozygous for these three pseudogene-like substitutions (arrows), but in fact their second vWF allele is a null allele (no mRNA). Family member in lane 3 is heterozygous for these three substitutions.

One hypothesis holds that gene conversion is a recombinational event and that it is regulated by special sites, the so-called *chi* sequence (5′-GCTGGTGG-3′) (31–33). An association between the presence of *chi*-like sequences and gene conversion events in humans has been suggested in

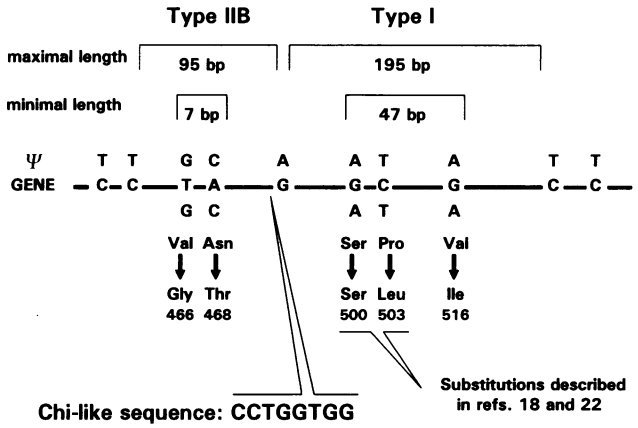

FIG. 4. Schematic representation of the 3′ end of intron 27 and the 5′ end of exon 28 of the vWF gene. Thick line represents vWF gene sequence; only the nucleotides at sites of mismatches with the pseudogene are indicated (GENE). Pseudogene sequence at these sites is also indicated (ψ). Multiple substitutions in the type IIB and type I families are shown. Minimal and maximal sizes of the putatively converted sequences are indicated, as are location of the *chi*-like sequence and multiple substitutions described by Holmberg *et al.* (18) and Zhang *et al.* (22).

steroid 21-hydroxylase deficiency (34) and for the β-globin gene (28). Therefore, we searched the vWF gene for the presence of *chi* sequences. In the gene sequence from exon 23 to 34, which corresponds to the pseudogene sequence, we have found four consensus *chi* sequences (5′-GCTGGTGG-3′). Two of these are in proximity of the putative gene conversions (one in intron 27 and one in exon 28). In addition, a *chi*-like sequence (5′-CCTGGTGG-3′; nucleotides 3726–3733) is present between the two gene conversions in exon 28 (Fig. 4). This sequence has one mismatch with the consensus *chi* sequence, but it has been reported that sequences similar but not identical to *chi* also stimulate recombination (32). The presence of *chi* sequences may be responsible for the relatively frequent conversion-like events in exon 28.

This work was supported by the Praeventiefonds (Grant 28-1862) and the NWO Foundation for Medical and Health Research Medigon (Grant 900-716-074).

1. Ruggeri, Z. M. & Zimmerman, T. S. (1987) *Blood* **70**, 895–904.
2. Hoyer, L. W. & Shainoff, J. R. (1980) *Blood* **55**, 1056–1059.
3. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* **264**, 19514–19527.
4. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Lester-Mancuso, T. L., Le Beau, M. M., Sorace, J. M. & Sadler, J. E. (1991) *Biochemistry* **30**, 253–269.
5. Ginsburg, D. & Sadler, J. E. (1993) *Thromb. Haemost.* **69**, 177–184.
6. Eikenboom, J. C. J., Reitsma, P. H., Peerlinck, K. M. J. & Briët, E. (1993) *Lancet* **341**, 982–986.
7. De Groot, P. G., Federici, A. B., De Boer, H. C., D'Alessio, P., Mannucci, P. M. & Sixma, J. J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 3793–3797.
8. Eikenboom, J. C. J., Ploos van Amstel, H. K., Reitsma, P. H. & Briët, E. (1992) *Thromb. Haemost.* **68**, 448–454.
9. Bonthron, D., Orr, E. C., Mitsock, L. M., Ginsburg, D., Handin, R. I. & Orkin, S. H. (1986) *Nucleic Acids Res.* **14**, 7125–7127.
10. Bachmann, B., Lüke, W. & Hunsmann, G. (1990) *Nucleic Acids Res.* **18**, 1309.
11. Vink, T. & Sixma, J. J. (1991) in *Biotechnology of Plasma Proteins*, eds. Albertini, A., Lenfant, C. L., Mannucci, P. M. & Sixma, J. J. (Karger, Basel), No. 58, pp. 78–83.
12. Cooney, K. A., Nichols, W. C., Bruck, M. E., Bahou, W. F., Shapiro, A. D., Bowie, E. J. W., Gralnick, H. R. & Ginsburg, D. (1991) *J. Clin. Invest.* **87**, 1227–1233.
13. Cooney, K. A., Lyons, S. E. & Ginsburg, D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2869–2872.
14. Randi, A. M., Rabinowitz, I., Mancuso, D. J., Mannucci, P. M. & Sadler, J. E. (1991) *J. Clin. Invest.* **87**, 1220–1226.
15. Lillicrap, D., Murray, E. W., Benford, K., Blanchette, V. S., Rivard, G. E., Wensley, R. & Giles, A. R. (1991) *Br. J. Haematol.* **79**, 612–617.
16. Donnér, M., Kristoffersson, A. C., Lenk, H., Scheibel, E., Dahlbäck, B., Nilsson, I. M. & Holmberg, L. (1992) *Br. J. Haematol.* **82**, 58–65.
17. Piétu, G., Ribba, A.-S., De Paillette, L., Chérel, G., Lavergne, J. M., Bahnak, B. R. & Meyer, D. (1992) *Blood Coag. Fibrinol.* **3**, 415–421.
18. Holmberg, L., Dent, J. A., Schneppenheim, R., Budde, U., Ware, J. & Ruggeri, Z. M. (1993) *J. Clin. Invest.* **91**, 2169–2177.
19. Zhang, Z. P., Lindstedt, M., Falk, G., Blombäck, M., Egberg, N. & Anvret, M. (1992) *Am. J. Hum. Genet.* **51**, 850–858.
20. Dónner, M., Kristoffersson, A.-C., Berntorp, E., Scheibel, E., Thorsen, S., Dahlbäck, B., Nilsson, I. M. & Holmberg, L. (1993) *Eur. J. Haematol.* **51**, 38–44.
21. Mancuso, D. J., Montgomery, R. R. & Adam, P. (1991) *Blood* **78** (Suppl 1), 67 (abstr.).
22. Zhang, Z. P., Blombäck, M., Nyman, D. & Anvret, M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7937–7940.
23. Kobayashi, I. (1992) *Adv. Biophys.* **28**, 81–133.
24. Powers, P. A. & Smithies, O. (1986) *Genetics* **112**, 343–358.
25. Baltimore, D. (1981) *Cell* **24**, 592–594.
26. Higashi, Y., Tanae, A., Inoue, H., Hiromasa, T. & Fujii-Kuriyama, Y. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7486–7490.
27. Gorski, J. & Mach, B. (1986) *Nature (London)* **322**, 67–70.
28. Matsumo, Y., Yamashiro, Y., Yamamoto, K., Hattori, Y., Yamamoto, K., Ohba, Y. & Miyaji, T. (1992) *Hum. Genet.* **88**, 357–358.
29. Hayashida, H., Kuma, K. & Miyata, T. (1992) *J. Mol. Evol.* **35**, 181–183.
30. Haber, J. E. (1992) *Curr. Opin. Cell. Biol.* **4**, 401–412.
31. Smith, G. R. (1983) *Cell* **34**, 709–710.
32. Cheng, K. C. & Smith, G. R. (1987) *J. Mol. Biol.* **194**, 747–750.
33. Kenter, A. L. & Birshtein, B. K. (1981) *Nature (London)* **293**, 402–404.
34. Amor, M., Parker, K. L., Globerman, H., New, M. I. & White, P. C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1600–1604.