Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 10730–10733, November 1993
Genetics

The classical human phosphoglucomutase (PGM1) isozyme polymorphism is generated by intragenic recombination

R. E. MARCH, W. PUTT, M. HOLLYOAKE, J. H. IVES, J. U. LOVEGROVE, D. A. HOPKINSON, Y. H. EDWARDS, AND D. B. WHITEHOUSE

Medical Research Council Human Biochemical Genetics Unit, Galton Laboratory, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, United Kingdom

Communicated by James V. Neel, June 17, 1993

ABSTRACT **The molecular basis of the classical human phosphoglucomutase 1 (PGM1) isozyme polymorphism has been established. In 1964, when this genetic polymorphism was first described, two common allelozymes PGM1 1 and PGM1 2 were identified by starch gel electrophoresis. The PGM1 2 isozyme showed a greater anodal electrophoretic mobility than PGM1 1. Subsequently, it was found that each of these allelozymes could be split, by isoelectric focusing, into two subtypes; the acidic isozymes were given the suffix + and the basic isozymes were given the suffix −. Hence, four genetically distinct isozymes 1+, 1−, 2+, and 2− were identified. We have now analyzed the whole of the coding region of the human *PGM1* gene by DNA sequencing in individuals of known PGM1 protein phenotype. Only two mutations have been found, both C to T transitions, at nt 723 and 1320. The mutation at position 723, which changes the amino acid sequence from Arg to Cys at residue 220, showed complete association with the PGM1 2/1 protein polymorphism: DNA from individuals showing the PGM1 1 isozyme carried the Arg codon CGT, whereas individuals showing the PGM1 2 isozyme carried the Cys codon TGT. Similarly, the mutation at position 1320, which leads to a Tyr to His substitution at residue 419, showed complete association with the PGM1+/− protein polymorphism: individuals with the + isozyme carried the Tyr codon TAT, whereas individuals with the − isozyme carried the His codon CAT. The charge changes predicted by these amino acid substitutions are entirely consistent with the charge intervals calculated from the isoelectric profiles of these four PGM1 isozymes. We therefore conclude that the mutations are solely responsible for the classical PGM1 protein polymorphism. Thus, our findings strongly support the view that only two point mutations are involved in the generation of the four common alleles and that one allele must have arisen by homologous intragenic recombination between these mutation sites.**

Genetic polymorphism is found in natural populations of all living things. Most protein polymorphisms appear to arise from single amino acid substitutions as a result of point mutations, approximately a third of which lead to substitution involving a charged amino acid that can be detected by electrophoresis (1). Other mechanisms underlying protein polymorphism have been identified, including mutations in noncoding sequence, small deletions and insertions, as well as more extensive rearrangements of structural genes. There are also instances of both intergenic and intragenic recombination events leading to protein variation. Well-known human examples include complete gene duplication by homologous nonreciprocal recombination, leading to gene copy number polymorphism in the α-globin locus (2) and recombination leading to partial gene duplication, which is the basis of the haptoglobin 2 allele (3). Similarly, unequal intragenic recombination can lead to the formation of hybrid genes and thus hybrid proteins, as is the case in the globin genes coding for Lepore hemoglobin (2), for example, the production of anomalous visual pigments from the recombination of red and green pigment genes (4), and length polymorphisms in proline-rich protein genes (5). In all of these cases, there is unequal (i.e., nonreciprocal) crossing-over, which inevitably leads to gain or loss of genetic material and a rather unusual protein variant. In contrast, intragenic reciprocal recombination leads to the exchange of genetic information without alteration in the overall size of the locus involved and relatively subtle protein variation, which may not be so easy to distinguish from the usual polymorphisms involving point mutations.

Several years ago, it was predicted from protein analysis that the classical human phosphoglucomutase (PGM1) isozyme polymorphism was attributable in part to intragenic reciprocal recombination (6–8). The protein shows a high level of heterozygosity in all human populations (9) and is an anchor point for linkage analysis on the short arm of chromosome 1. The 10 common phenotypes are encoded by four alleles, *PGM1*1+*, *1−*, *2+*, and *2−* (10). It has been suggested that one of these alleles is generated by intragenic recombination in between the sites of the mutations underlying the 2/1 and +/− features of the PGM1 isozymes (6). This hypothesis was based on the observation that the difference in the isoelectric points (approximately pH 0.1) between the + and − features is similar for both the 1 and 2 allelomorphs. Thus, if it is assumed that *1+* is the ancestral allele and that the − and the *2* features have arisen by point mutations, then intragenic recombination would produce the *2−* allele (6). Four other less common variants, *PGM1*7+*, *7−*, *3+*, and *3−*, have been included in this hypothesis by postulating a further single point mutation and three intragenic recombination events (7, 8).

In a recent study, we have provided indirect evidence for crossing-over between the sites of the putative *2/1* and +/− mutations from analysis of a polymorphic site in the 3′ untranslated region of the *PGM1* gene, since we observed allelic association of the 3′ polymorphic site with the +/− phenotypes but not with the 2/1 phenotypes (11). We now describe the direct investigation of mutations in the coding region of the *PGM1* gene in individuals of known protein phenotype. We have identified two point mutations only, both of which lead to amino acid substitutions and these correlate exactly with the 2/1 and +/− phenotypes. Thus, we have obtained direct confirmation for the hypothesis that intragenic recombination underlies the classical PGM1 isozyme polymorphism.

MATERIALS AND METHODS

Samples of Known PGM1 Protein Phenotype. A representative panel of nine individuals was chosen so as to include 7 of

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.

Abbreviation: SSCP, single-strand conformation polymorphism.

the 10 PGM1 isozyme phenotypes: 1+ (n = 2), 1−, 1+1−, 2+ (n = 2), 2−, 2+2−, and 2−1− by screening random blood samples using isoelectric focusing with gradient pH 5–7 (12). DNA was extracted from 6 ml of whole blood on an Applied Biosystems nucleic acid extractor (model 340A) using an automated procedure recommended by the manufacturer. In addition, crude preparations of RNA were made from six lymphoblastoid cell lines of PGM1 phenotypes 1+, 1−, 1+1−, 2+2−, 2+1+, and 2−1−, by lysis of 10^6 cells in 0.1% Nonidet P-40 and extraction with phenol/chloroform.

Search for Mutations. Two approaches were used to screen the *PGM1* coding sequence for the presence of point mutations. We carried out reverse transcriptase/PCR with total RNA prepared from lymphoblastoid cell lines of known PGM1 phenotype and determined the DNA sequence of the PCR product. We also performed sequence analysis of all 11 *PGM1* exons by PCR from genomic DNA.

Amplification from RNA. The primers used to amplify cDNA were derived from the published *PGM1* cDNA sequence (13). Ten micrograms of total RNA was reverse transcribed in the presence of 100 pmol of reverse primer, 200 units of reverse transcriptase (GIBCO/BRL), 0.2 mM dNTPs, and 1.5 mM $MgCl_2$ for 1 hr at 42°C. The cDNA/RNA product was amplified by addition of 2 units of *Taq* polymerase (Promega or Applied Biotechnology) and 50 pmol of forward primer in a final vol of 100 μl of PCR buffer. After 5 min of denaturation at 94°C, amplification of the cDNA template was carried out for 50 cycles at 94°C for 20 sec, 55°C for 45 sec, and 72°C for 45 sec. Ramps of 1.5–2.5 sec/°C were set before the annealing step and the extension step (14). Five microliters of the resulting PCR product was then used as the template in a second round (50 cycles) of amplification with primers nested 10–20 bases inside the original primers and with the forward primer biotinylated at the 5′ end.

Amplification of genomic DNA. The primers used in the amplification of genomic DNA were derived from human *PGM1* intron sequence (15). Those either side of exons 4 and 8 were as follows: 4F, GCAGGTTTACAGCAATATAGTCACA; 4R, TGAAGCATCATGATACACACAGAAG; 8F, GGGATGCAGAGCCAAACCATATCAAG; 8R, TAAGACAGGAGAGGCTGTGGATGCG. Five hundred nanograms of genomic DNA was amplified in the presence of 50 pmol of forward and reverse primers, 2 units of *Taq* polymerase, 2 mM dNTPs, and 1.5 mM $MgCl_2$. The forward primer was biotinylated at the 5′ end. Fifty cycles of amplification were carried out at 94°C for 20 sec, average t_m −10°C for 45 sec, and 72°C for 45 sec.

DNA Sequence Analysis. Single-stranded DNA templates for sequence analysis were prepared in two ways. PCR products in which one strand was biotinylated at the 5′ end were separated into single strands with streptavidin-coated magnetic beads (Dynal). Alternatively, PCR products were ligated into M13 using a T-vector cloning system (16) and single-stranded DNA was prepared by standard methods. Sequence analysis was carried out using the Sequenase system (United States Biochemical) by the dideoxynucleotide chain-termination method (17).

***PGM1* Gene Polymorphisms.** DNA was prepared by PCR from exons 4 and 8 from 65 unrelated individuals, including the representative panel mentioned above, for single-strand conformation polymorphism (SSCP) analysis by procedures already used to examine the 3′ region of *PGM1* (11). Rapid flat-bed electrophoresis of PCR products was carried out on native 20% polyacrylamide gels using the automated Phastsystem (Pharmacia LKB) at 5°C for the exon 4 SSCP and 10°C for the exon 8 SSCP. These PCR products were further analyzed by digestion with restriction endonucleases *Bgl* II (GIBCO/BRL), *Alw* I, and *Nla* III (New England Biolabs) according to the manufacturer's recommendations and agarose gel electrophoresis.

RESULTS

The Molecular Basis of the Common Protein Polymorphism. Only two mutations were encountered in sequencing the entire coding region; the complete DNA sequence was analyzed in four individuals (PGM1 phenotype 1+, 1−, 2+, 2+2−) and specific exons (4 and 8) were analyzed in a further 10 individuals. The two mutations were a C to T transition in exon 4 at nt 723 in the mRNA sequence and a C to T transition in exon 8 at nt 1320.

The mutation at nt 723 showed complete association with the PGM1 2/1 protein polymorphism (Fig. 1*A*). All individuals who were homozygous for the isozyme phenotype PGM1 1 carried the base C, whereas individuals who were homozygous for PGM1 2 carried the base T. Heterozygous individuals showed both bases comigrating at this position with the bands at approximately half the intensity of the neighboring bands. Similarly, the mutation at nt 1320 showed complete association with the PGM1 +/− protein polymorphism (Fig. 1*B*). All individuals who were homozygous for the + protein phenotype carried the base T, whereas individuals who were homozygous for the − protein phenotype carried the base C. Individuals who were +/− heterozygotes showed both bases at this position. Therefore, we conclude that the C to T transitions at nt 723 and 1320 constitute the molecular basis of the classical PGM1 protein polymorphism. These findings indicate that only two point mutations are involved in generation of the four common allelomorphs and strongly support the view that one of these has arisen by

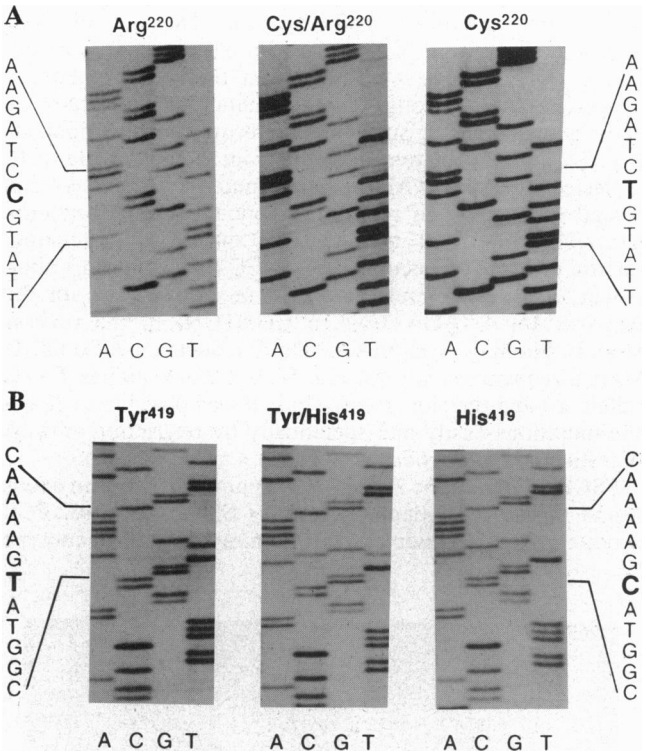

FIG. 1. (*A*) Sequencing gel of exon 4 PCR products from three individuals of PGM1 protein phenotype 1−, 2−1−, and 2+, showing C or T at nt 723 in samples with the 1 or 2 phenotype, respectively, and the associated Arg to Cys substitution at residue 220. The sample from the individual heterozygous for the 2/1 protein phenotype shows both a C and a T band comigrating at this position. (*B*) Sequencing gel of exon 8 PCR products from three individuals of PGM1 protein phenotype 2+, 2+2−, and 2−1− showing T or C at position 1320 in samples with the + and − phenotype, respectively, and the associated Tyr to His substitution at residue 419. The sample from the individual heterozygous for the +/− protein phenotype shows both a C and a T band comigrating at this position.

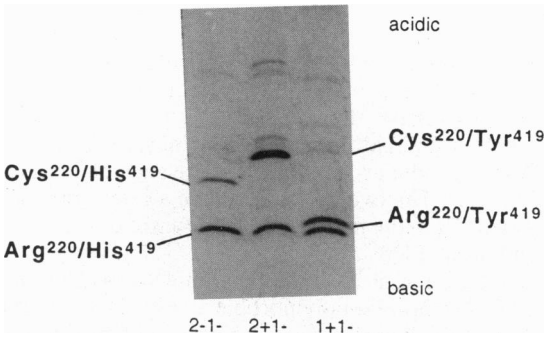

FIG. 2. Immunoblot of isoelectric focusing gel of placental samples of different PGM1 phenotypes showing the deduced amino acid composition at residues 220 and 419 in the four common isozymes (1+, 1−, 2+, and 2−).

reciprocal intragenic recombination at a point between the two sites of mutation.

The C to T transition at nt 723 (exon 4) changes the amino acid sequence from an Arg to a Cys at residue 220. This basic to neutral amino acid substitution is entirely consistent with the more anodal electrophoretic property of the PGM1 2 isozyme compared with the PGM1 1 isozyme. Similarly, the transition at nt 1320 (exon 8) leads to a Tyr to His substitution at residue 419. This rather slight neutral to weak basic amino acid change is consistent with the relatively small charge difference seen between the + and the − isozyme band recognized on isoelectric focusing (Fig. 2). The three-dimensional structure of the PGM1 protein has recently been determined (18) and is shown to be organized in four domains. As might be expected from their electrophoretic properties, the positions of the two mutations at residues 220 and 419 map near the surface of the protein, within domains 2 and 3, and are relatively remote from the active site cleft.

Restriction Enzyme Analysis. The mutations in exons 4 and 8 lead to changes in restriction endonuclease recognition sites. Thus in exon 4, the AGATCT site for *Bgl* II (characteristic of allele *2*) becomes AGATCC (allele *1*), thus abolishing this site but creating a new recognition site for the enzyme *Alw* I (AGATCN$_4$ to GGATCN$_4$ in the reverse strand). Similarly in exon 8 the DNA sequence CATG (allele −) is a recognition site for *Nla* III but this becomes TATG (allele +) and the site is lost. Thus, it was possible to detect the mutations easily and specifically by restriction enzyme digestion of PCR products of exons 4 and 8 (Fig. 3).

SSCP Analysis. The *2/1* and +/− point mutations in exons 4 and 8 were also demonstrable as SSCPs (Fig. 4). PCR products from 65 individuals of known PGM1 phenotype

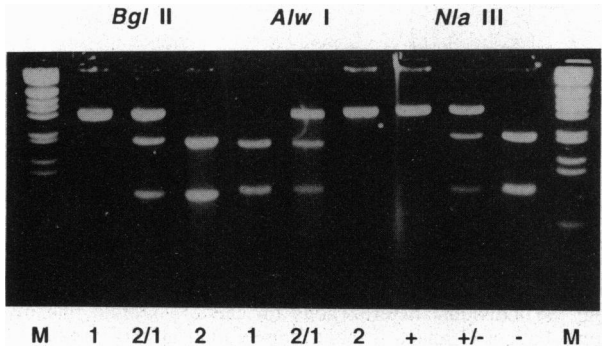

FIG. 3. Agarose gel electrophoresis of PCR products from exons 4 and 8 from individuals heterozygous or homozygous for the 2/1 and +/− protein polymorphism digested with *Bgl* II and *Alw* I in the case of exon 4 (*2/1*) and *Nla* III for exon 8 (+/−). Lane M, size markers (GIBCO/BRL; 1-kb ladder).

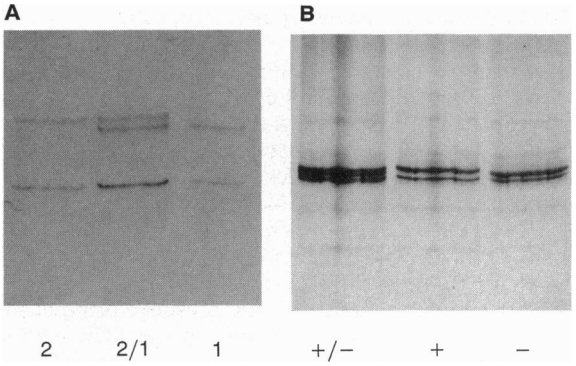

FIG. 4. SSCP analysis of exon 4 PCR products from three individuals of PGM1 protein phenotype 2, 2/1, and 1 (full subtypes, 2+, 2−1−, and 1−) (*A*) and exon 8 PCR products from three individuals of PGM1 protein phenotype +/−, +, and − (full subtypes, 2+2−, 2+, and 2−1−) (*B*).

were examined. The *2/1* mutations were well resolved, with each of the homozygotes showing a distinct two-banded pattern and the heterozygote showing a three-banded pattern; the mobility difference between the + and − mutations was less marked, but each of the homozygotes (two banded) and the heterozygote (four banded) were clearly distinguishable. A single exceptional pattern was found to reflect heterozygosity for a point mutation in intron sequence flanking exon 8.

DISCUSSION

We have shown beyond doubt that the electrophoretic differences between the four common PGM1 isozyme alleles are due to four combinations of two point mutations that underlie the 2/1 and +/− isozyme features. Thus, at the DNA level the structural basis of the isozyme alleles is determined by four haplotypes; *1+*, *1−*, *2+*, and *2−*.

These findings strongly support the hypothesis (6, 7) that the common PGM1 polymorphism has involved intragenic recombination between two point mutations. The view is further supported by two independent studies of PGM1 isozyme phenotypes in a total of ≈20,000 mother/child pairs from the Northern European population, which have found convincing evidence for reciprocal recombination within the *PGM1* gene (ref. 19; J. Dissing, personal communication). Both studies report cases in which neither of the maternal PGM1 isozyme configurations was transmitted to the child. In every one of these cases, the mother was a double heterozygote, such as *2+/1−*, from whom the child had received a recombined haplotype—for example, either *2−* or *1+*. Thus, the new maternally derived haplotype appeared to be a product of germ-line intragenic recombination between the *2/1* and +/− sites. Wetterling (19) estimated the overall recombination rate within the *PGM1* locus to be on the order of 0.05%. A second line of evidence that indirectly supports the intragenic recombination hypothesis can be derived from allelic association studies between the 2/1 and +/− protein features using data from 13 samples (n = 140–12,000) from major ethnic groups (20, 21) and Ott's associate program (22). Significant linkage disequilibrium (Δ) was found in only four populations and the median value of Δ/D_{max} for all populations was only 0.10 (range, 0.014–0.523).

Thus, the nucleotide sequence analysis of the *PGM1* common alleles, the mother/child studies, and the allele association analysis point to the existence of a node where reciprocal intragenic recombination occurs, at high frequency, between the *2/1* and +/− mutation sites. The mutation sites lie in exons 4 (*2/1*) and 8 (+/−) of the human *PGM1* gene. Our recent elucidation of the genomic structure

of *PGM1* shows that they are separated by 18 kb of DNA (15). This is a relatively short distance and a precise localization of the site of recombination should be possible from detailed analysis of haplotypes covering this region. We have already characterized five polymorphic sites in the *PGM1* gene: the *2/1* site, +/− site, 3′ site (11), and two *Taq* I restriction fragment length polymorphisms (23).

We cannot be certain of the true ancestor haplotype from human data, although on the basis of its high frequency in the majority of populations, the *1+* arrangement is the most obvious candidate. In this context, it is interesting that the rabbit PGM1 amino acid sequence reported from three independent sources (13, 24, 25) and the mouse homologue (Pgm-2) deduced amino acid sequence (J. Friedman, personal communication) all have an Arg at position 220. Thus, they are all *1*-like, which establishes Arg-220 as an ancient character predating the evolutionary events leading to the separation of the lines for rodents, lagomorphs, and primates. However, neither rabbit nor mouse has His or Tyr at position 419; instead, both have a Phe. Thus, at the level of expressed variation, the appearance of the + or − isozyme feature is relatively recent, perhaps occurring during primate evolution (6). At present, we have no information on the evolutionary age of the recombination site but it is tempting to speculate that, like Arg-220, it too may be ancient and conserved throughout the animal and plant kingdoms. This could account, in part, for the high incidence of PGM polymorphism found in most species.

In more general terms, these studies suggest that the importance of intragenic recombination has probably been underestimated and neglected as an evolutionary force for generating expressed diversity. It is possible that closer scrutiny of other common protein polymorphisms at the DNA sequence level will show that some of the variation previously ascribed solely to point mutation in fact arises also by reciprocal intragenic recombination. The human *PGM1* locus provides an ideal model system for commencing the analysis of this facet of genetic polymorphism.

We thank Steve Jeremiah for preparation of DNA samples. R.E.M. and M.H. are supported by a grant from the Home Office.

1. Harris, H. & Hopkinson, D. A. (1976) *Handbook of Enzyme Electrophoresis in Human Genetics* (North-Holland, Amsterdam).
2. Weatherall, D. J., Clegg, J. B., Higgs, D. R. & Wood, W. G. (1989) in *The Metabolic Basis of Inherited Disease*, eds. Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (McGraw-Hill, New York), 6th Ed., pp. 2281–2339.
3. Bowman, B. H. & Kurosky, A. (1982) *Adv. Hum. Genet.* **12,** 189–261.
4. Nathans, J., Merbs, S. L., Sung, C.-H., Weitz, C. J. & Wang, Y. (1992) *Annu. Rev. Genet.* **26,** 403–424.
5. Lyons, K. M., Stein, J. H. & Smithies, O. (1988) *Genetics* **120,** 267–278.
6. Carter, N. D., West, C. M., Emes, E., Parkin, B. & Marshall, W. H. (1979) *Ann. Hum. Biol.* **6,** 221–230.
7. Takahashi, N., Neel, J. V., Satoh, C., Nishizaki, J. & Masunari, N. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6636–6640.
8. Neel, J. V., Satoh, C., Smouse, P., Asakawa, J., Takahashi, N., Goriki, K., Fujita, M., Kageoka, T. & Hazama, R. (1988) *Am. J. Hum. Genet.* **43,** 870–893.
9. Roychoudhury, A. K. & Nei, M. (1988) *Human Polymorphic Genes World Distribution* (Oxford Univ. Press, New York), pp. 109–110.
10. Bark, J. E., Harris, M. J. & Firth, M. (1976) *J. Forensic Sci. Soc.* **16,** 115–120.
11. March, R. E., Hollyoake, M., Putt, W., Hopkinson, D. A., Edwards, Y. H. & Whitehouse, D. B. (1993) *Ann. Hum. Genet.* **57,** 1–8.
12. Drago, G. A., Hopkinson, D. A., Westwood, S. A. & Whitehouse, D. B. (1991) *Ann. Hum. Genet.* **55,** 263–271.
13. Whitehouse, D. B., Putt, W., Lovegrove, J. U., Morrison, K., Hollyoake, M., Fox, M. F., Hopkinson, D. A. & Edwards, Y. H. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 411–415.
14. Kawasaki, E. S. (1990) in *PCR Protocols*, eds. Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (Academic, New York), pp. 21–27.
15. Putt, W., Ives, J. H., Hollyoake, M., Hopkinson, D. A., Whitehouse, D. B. & Edwards, Y. H. (1993) *Biochem. J.* **296,** in press.
16. Marchuk, D., Drumm, M., Saulino, A. & Collins, F. S. (1990) *Nucleic Acids Res.* **19,** 1154.
17. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Dai, J. B., Liu, Y., Ray, W. J., Jr., & Konno, M. (1992) *J. Biol. Chem.* **267,** 6322–6337.
19. Wetterling, G. (1990) *Adv. Forensic Haemogenet.* **3,** 218–221.
20. Stedman, R. & Rothwell, T. J. (1985) *J. Forensic Sci. Soc.* **25,** 95–144.
21. Ruofu, D., Zhi, Z. & Hong, Z. (1992) *Gene Geography* **6,** 21–26.
22. Ott, J. (1985) *Genet. Epidemiol.* **2,** 79–84.
23. Hollyoake, M., Putt, W., Edwards, Y. H. & Whitehouse, D. B. (1992) *Hum. Mol. Genet.* **1,** 354.
24. Ray, W. J., Jr., Hermodson, M. A., Jr., Puvathingal, J. M. & Mahoney, W. C. (1983) *J. Biol. Chem.* **258,** 9166–9174.
25. Lee, Y. S., Marks, A. R., Gureckas, N., Lacro, R., Nadal-Ginard, B. & Kim, D. H. (1992) *J. Biol. Chem.* **267,** 21080–21088.