

Phosphoglucosmutase 1: Complete human and rabbit mRNA sequences and direct mapping of this highly polymorphic marker on human chromosome 1

D. B. WHITEHOUSE, W. PUTT, J. U. LOVEGROVE, K. MORRISON, M. HOLLYOAKE, M. F. FOX,
D. A. HOPKINSON, AND Y. H. EDWARDS

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, September 23, 1991

ABSTRACT A cDNA clone encoding the mRNA for the highly polymorphic human enzyme phosphoglucosmutase 1 (PGM1; EC 5.4.2.2) has been isolated and characterized. This was achieved indirectly by first isolating a rabbit cDNA from an expression library using anti-rabbit PGM antibodies. A comparison of the nucleotide sequences shows that the homologies between human and rabbit PGM1 mRNAs are 92% and 97% for the coding nucleotide sequence and the amino acid sequence, respectively. The derived rabbit amino acid sequence is in complete agreement with the published protein sequence for rabbit muscle PGM. A physical localization of the human PGM1 gene to chromosome 1p31 has been determined by *in situ* hybridization. Analysis of DNA from a wide variety of vertebrates indicates a high level of PGM1 sequence conservation during evolution.

Phosphoglucosmutase (PGM; EC 5.4.2.2) catalyzes the interconversion of glucose 1-phosphate and glucose 6-phosphate and thus has a pivotal role in glycolysis and gluconeogenesis. The human enzyme is encoded by three autosomal unlinked structural loci, designated PGM1, PGM2, and PGM3 (1, 2), and there is evidence for a fourth locus expressed in human milk (3). On the basis of protein studies, it is assumed that the loci arose by gene duplication and evolved separately to provide the present day isozymes, showing distinctive properties, such as molecular size (4), thermostability (5), substrate specificity (6), tissue distribution (7), and sulfhydryl group reactivity (8). Multiple isozymes, attributable to multiple PGM loci have been identified in a very wide range of species and it is evident that the initial PGM gene duplication(s) occurred in the remote evolutionary past.

Isozyme polymorphism of human PGM1 was identified more than 25 years ago (9) and the locus has since been shown to exhibit a very high incidence of both common and rare allelic variation in all populations (10). The PGM1 locus is an important anchor point for linkage analysis and for positioning markers on human chromosome 1p and its homologues in other species. The heterozygosity demonstrable by isozyme analysis of human erythrocytes [0.81 in the Centre d'Etude Polymorphisme Humain (CEPH) family panel for example] is comparable to several of the variable number tandem repeat loci and exceeds most of the restriction fragment length polymorphisms recognized by other probes on chromosome 1. The PGM1 isozyme polymorphism is a particularly useful marker of individuality in forensic science (11, 12) since the discriminating power of the PGM1 locus (1 – probability of a match) is about 0.75 for the common isozyme phenotypes revealed by isoelectric focusing (12). Furthermore, the isozymes are found in a variety of body fluids and enzyme

activity is retained for prolonged periods in stains recovered from scenes of crimes.

The extensive heterozygosity of PGM1 and certain features of the variant isozyme patterns have led to speculation about the role of intragenic recombination in generating this diversity. The isozymes associated with the two alleles, PGM1*1 and PGM1*2, discovered by starch gel electrophoresis, can be subdivided by isoelectric focusing into two classes: one designated “+” is more anodal than the other, designated “–”. Thus four alleles (PGM1*1⁺, 1[–], 2⁺, and 2[–]) define 10 common phenotypes (11). Furthermore, the properties of these common alleles 1 vs. 2 and + vs. – do not occur in a random association. Similarly, the isozymes encoded by the less-common PGM alleles PGM1*3 and PGM1*7 can also be subdivided by isoelectric focusing into + and – categories and here also the association of the latter property with the electromorphs (PGM1*3 vs. PGM1*7) is not random (13). These findings have led to the suggestion that much of the PGM1 diversity might be attributed to intragenic recombination. A gene phylogeny has been constructed for the eight most common alleles, in which intragenic recombination plays at least as significant a role as nucleotide substitution in the generation of allelic diversity (13).

In the present communication, we describe the isolation and the full sequence of both human and rabbit PGM1 mRNA. Antibodies raised in a sheep against purified rabbit muscle PGM cross react with human PGM1 (14) but are not sufficiently sensitive to detect human PGM1 expressed in phage expression vectors. Thus we approached the cloning of human PGM1 by using the antibodies to isolate a rabbit PGM1 cDNA clone, which was then employed as a human library screening probe. The cloning of the human PGM1 cDNA[†] will allow the hypothesis of the origin of the allelic diversity to be fully tested.

MATERIALS AND METHODS

Cloning of Human and Rabbit PGM1. Poly(A)⁺ RNA was isolated from rabbit hind leg muscle by selective LiCl precipitation and binding to oligo(dT)-cellulose. A cDNA library (2 × 10⁶ recombinants) was produced in the expression vector λgt11 using standard procedures with EcoRI adapters (15) and screened using antibodies raised in a sheep and specific for rabbit muscle PGM1 (14). Positive plaques were recloned three times to yield plaque-purified phage. cDNA inserts were removed by EcoRI digestion, gel-purified, and used to screen a human muscle cDNA library (15). The human PGM1 recombinants were similarly plaque-purified.

Abbreviations: PGM, phosphoglucosmutase; CEPH, Centre d'Etude Polymorphisme Humain.

[†]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M83088).

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.

Large-scale preparations of purified λ recombinants were carried out using a standard cesium chloride centrifugation procedure (16).

Subcloning and Labeling of cDNA Probes. Both human and rabbit cDNA inserts were subcloned into *EcoRI* sites of pUC18 and M13mp18 vectors. Preparations of plasmid or the replicative form of M13 were purified by equilibrium centrifugation in cesium chloride (16). cDNA inserts from pUC18 recombinants were gel-purified and labeled with ^{32}P using the oligonucleotide primer method (17).

DNA Sequence Analysis. M13mp18 subclones containing PGM1 cDNA inserts in opposite orientations were identified by restriction site mapping. Both cDNAs were sequenced by the dideoxynucleotide chain-termination sequencing method of Sanger *et al.* (18). For rabbit PGM1, intact cDNA and various fragments obtained by restriction enzyme digestion with *Kpn* I, *Hind*III, or *Sph* I were sequenced in both directions and, for the human PGM1, several *Hinf*I fragments were sequenced in both directions and a series of overlapping bidirectional deletions using exonuclease III (19) were created and sequenced.

Nucleic Acid Gel Blot Analysis. For Southern blots, genomic DNA was extracted, digested, electrophoresed on 0.8% agarose, and transferred to Hybond N⁺ (Amersham). Total and poly(A)⁺ RNA was purified from cell cultures or tissue samples fractionated by 3-(*N*-morpholino)propane-sulfonic acid (Mops)/formaldehyde electrophoresis (16) and blotted onto GeneScreen membrane (NEN). Insert cDNA (25 ng) was labeled using [^{32}P]dCTP to a specific activity of 1×10^8 cpm/ μg for both Northern and Southern blots. Southern blots were hybridized in 10% (wt/vol) dextran sulfate/5 \times standard saline citrate (SSC)/0.1% SDS at 65°C and Northern blots were hybridized in 50% (vol/vol) formamide/5 \times SSPE (1 \times SSPE = 0.18 M NaCl/10 mM sodium phosphate, pH 7.4/1 mM EDTA) at 42°C. Membranes were washed in 1 \times SSC/0.1% SDS at 65°C for Southern blots and in 0.2 \times SSC/0.1% SDS at 50°C for RNA analysis. Autoradiography was at -70°C.

In Situ Hybridization. *In situ* hybridization was carried out on spreads of human metaphase chromosomes by using a biotinylated whole PGM1 cosmid as probe as described (20) but with certain modifications. Human Cot₁ DNA, a fraction of total DNA in which highly repetitive sequences are prevalent, was used at a 50-fold excess of probe concentration to suppress hybridization to these sequences. Signal detection was achieved using fluorescein isothiocyanate-conjugated avidin and amplified by alternate applications of biotinylated anti-avidin and fluorescein isothiocyanate-conjugated avidin. Preparations were mounted in anti-fade solution *p*-phenylenediamine dihydrochloride to which the fluorochromes diamidinophenylindole and propidium iodide had been added for counterstaining and banding. Preparations were evaluated using epifluorescence microscopy. Separate images were obtained for each fluorochrome, collected electronically by laser scanning (Bio-Rad MRC 600 confocal microscope), and merged. Signal position was compared with R-banding patterns for at least 10 metaphases showing signal on both chromatids with negligible background.

PCR Analysis of Somatic Cell Hybrids. Details of the somatic cell hybrids used are listed either in Carritt *et al.* (21) or Wong *et al.* (22). Hybrid DNA or human or rodent parent DNA (3 μg) was amplified using 25-mer oligonucleotides from the 3' untranslated region of human PGM1 (see Fig. 1) as primers. After 30 cycles of amplification, each consisting of 15 s at 94°C, 30 s at 58°C, and 30 s at 72°C, 10 μl was removed and analyzed on a 2.0% agarose gel. Each hybrid DNA was also amplified with primers to rodent Trp53 pseudogene (C. Abbott, personal communication) as a control to establish that all samples were amplifiable.

RESULTS

cDNA Clones for Human PGM1. Two human PGM1 recombinants, λ HPGM1.1 and λ HPGM1.2, appeared to contain full-length cDNA inserts and detected an mRNA of ≈ 2.3 kilobases after Northern blot analysis of human skeletal muscle RNA. The length of the human cDNA inserts was 2320 base pairs (bp) and the length of the rabbit insert (from clone λ RPGM) was 2279 bp. Each contained an open reading frame of 1686 bp; when translated there was 97% homology between the human and rabbit derived amino acid sequences (Fig. 1) and complete agreement between the derived rabbit amino acid sequence and the published rabbit protein sequence for muscle PGM1 (23). The positions of the 18 amino acid substitutions found between the human and rabbit PGM1 are shown diagrammatically in Fig. 2. A region of 83 amino acids spanning the active site is completely conserved, with serine at position 116. Ser-116 is known to react in a phosphorylated form directly with the substrate glucose 1-phosphate (24). There is complete conservation between residues 220 and 417 except for a single amino acid difference at position 345 where serine is present in the human protein and asparagine is in the rabbit. In contrast, the COOH-terminal region is less well conserved with nine substitutions within the final 145 residues. The amino acid differences between the species account for a relatively high positive charge on the rabbit protein, which is reflected in the more basic isoelectric point (14). The molecular weights derived from the rabbit and human amino acid sequences are 61,600 (23) and 61,300, respectively; the latter is somewhat larger than the crude estimate of 51,000 obtained by gel-filtration chromatography of human placental PGM1 (4).

The cDNAs include a 5' untranslated region of 62 bp and 39 bp and a 3' untranslated region of 572 bp and 554 bp in the human and rabbit clones, respectively. The overall homology between the human and rabbit coding sequences is 92%. There is a moderate degree of homology at the 5' untranslated region which is G+C-rich (G+C = 76% and 80%, respectively) in both species; without sequence analysis or methylation studies of the 5' flanking region, it is not certain whether this represents a Hpa II tiny fragments (HTF) island, but such a feature might be expected in a ubiquitously expressed gene like PGM1. The sequence homology between the 3' untranslated regions is generally less but there are several segments extending over at least 10 bp of complete identity. Such patterns of homology in 3' untranslated regions have been encountered in other genes (25, 26) but their general significance is not known.

A Southern blot analysis of DNA from animal species indicates that PGM1 is conserved among a wide variety of vertebrates ranging from primates to birds and amphibia (Fig. 3).

PGM1 on Human Chromosome 1. Human PGM1 was assigned to chromosome 1 by isozyme analysis in rodent-human somatic cell hybrids (27) and the regional localization has been refined by linkage analysis to 1p22 (28). PCR analysis using primer oligonucleotides from the 3' untranslated region of the PGM1 cDNA confirms the assignment to 1p. Fig. 4 *Upper* shows the amplification of DNA from two hybrids, CRAB5 and CRAB4, which differ only by the presence of human chromosomes 1 and 3 in CRAB5 and their absence in CRAB4. Also shown are amplifications of DNA from TWIN19F9 and TWIN19F6, which differ only by the absence of chromosome 1 from TWIN19F6. (CRAB11, which is also shown in Fig. 4, is negative for chromosome 1 markers.) Fig. 4 also includes two hybrids HCH5 and HCH7 (20), which contain the reciprocal halves of the translocation t(1;10)(p36.31;q22.1). HCH7, which contains 1p36.1 to 1qter, is positive for the PGM1-amplified product whereas its counterpart HCH5 (1p36.1 to 1pter) is negative.

Hu		gg ggcggccgccccctcgcgacccaagtccgc	cgctctgacccccggcagcaagtgcgccacc	MetValLysIleValThrValLysThrGln	ATGGTGAAGATCGTGACAGTTAAGACCCAG	92
Ra		gaggatccg	ggtaccatggccccggcagcccgccacc		C G A	
	AlaTyrGlnAspGlnLysProGlyThrSer	GlyLeuArgLysArgValLysValPheGln	SerSerAlaAsnTyrAlaGluAsnPheIle	GlnSerIleIleSerThrValGluProAla	CAGAGTATCATCTCCAACCGTGGAGCCGGCG	212
	CGCTACCCAGGACCCAGAACCGGGCACGAGC	GGGCTGCGGAAGCGGGTGAAGGTGTCCAG	AGCAGCCCAACTACCGCGAGAACTTCATC			
		C	A	T		
	GlnArgGlnGluAlaThrLeuValValGly	GlyAspGlyArgPheTyrMetLysGluAla	IleGlnLeuIleAlaArgIleAlaAlaAla	AsnGlyIleGlyArgLeuValIleGlyGln	AACCGGATCGGTCCGCTTGTTCGGACAG	332
	CAGCGCGCAGGAGCCACCGTGGTGGGGC	GGGGACCGCCGGTCTCATCATGAAGAGGCC	ATCCAGCTCATCGCTCGCATCGCTGCCGCC			
	C	C	A	C	T G	
	AsnGlyIleLeuSerThrProAlaValSer	CysIleIleArgLysIleLysAlaIleGly	GlyIleIleLeuThrAlaSerHisAsnPro	GlyGlyProAsnGlyAspPheGlyIleLys	GGGGCCCAATGGAGATTTTGGATCAA	452
	AATGGAACTCTCCACCCCTGCTGTATCC	TGCATCATTTAGAAAATCAAAGCCATTTGGT	GGGATCATTTCTGACAGCCAGTCAACAACA			
		G	C	A	C A G	
	PheAsnIleSerAsnGlyGlyProAlaPro	GluAlaIleThrAspLysIlePheGlnIle	SerLysThrIleGluGluTyrAlaValCys	ProAspLeuLysValAspLeuGlyValLeu	CCTGACCTGAAGTAGACCTTGGTGTCTCG	572
	TTCAATATTTCTAATGGAGTCTCTGCTCCA	GAAGCAATAACTGATAAAATTTTCCAATTC	AGCAAGCAATTAAGAATATGCAATTTGTC			
	T	C		GA		
	GlyLysGlnGlnPheAspLeuGluAsnLys	PheLysProPheThrValGluIleValAsp	SerValGluAlaTyrAlaThrMetLeuArg	SerIlePheAspPheSerAlaLeuLysGlu	AGCATCTTTGATTTCACTGCATGAAGAA	692
	GGAAAGCAGCAGTTTTCCTTGGAAAATAAG	TTCAAACCCCTTCACTGGAAAATTTGGAT	TCGGTGAAGCTTATGCTCAATGCTGAGA			
	G	C	A	G	A A T	
	LeuLeuSerGlyProAsnArgLeuLysIle	CysIleAspAlaMetHisGlyValValGly	ProTyrValLysLysIleLeuCysGluGlu	LeuGlyAlaProAlaAsnSerAlaValAsn	CTCGGTGCCCTGCGAACTCGGCAGTTAAC	812
	CTACTTTCTGGGCCAAACCGACTGAAGATC	TGTATTGATGCTATGCTATGGAGTTTGGGA	CGTATGTAAAGAAGATCTCTGTGAAGAA			
	G C	A C	C	C	A A C T C	
	CysValProLeuGluAspPheGlyGlyHis	HisProAspProAsnLeuThrTyrAlaAla	AspLeuValGluThrMetLysSerGlyGlu	HisAspPheGlyAlaAlaPheAspGlyAsp	CATGATTTTGGGGCTGCTTGTATGGAGAT	932
	TGCGTCCCTTGGAGGACTTTGGAGCCCA	CACCCGTGACCCCACTCACCTATGCGAGCT	GACCTGGTGGAGCACATGAATCAGGAGAG			
	T	C G	C	C	C C	
	GlyAspArgAsnMetIleLeuGlyLysHis	GlyPhePheValAsnProSerAspSerVal	AlaValIleAlaAlaAsnIlePheSerIle	ProTyrPheGlnGlnThrGlyValArgGly	CCGTATTTCCAGCAGACGGGGTCCGGCC	1052
	GGGATCGAAAACATGATTTCTGGCAAGCAT	GGGTTCCTTTGTGAACCCCTCAGACTCTGTG	GCTGTATGCTGCTGCCAACATCTTCAGCATT			
	G	C	C	A	C	
	PheAlaArgSerMetProThrSerGlyAla	LeuAspArgValAlaSerAlaThrLysIle	AlaLeuTyrGluThrProThrGlyTrpLys	PhePheGlyAsnLeuMetAspAlaSerLys	TTTTTTGGGAATTTGATGGACCGCAGCAA	1172
	TTTGCACGAGCATGCCCAAGATGGTGTCT	CTGGACCGGGTGGCTAGTGTCTCAAAGATT	GCMTTGTATGAGACCCCACTGGCTGGAG			
	C	C	A	G A C C	T	T A
	LeuSerLeuCysGlyGluGluSerPheGly	ThrGlySerAspHisIleArgGluLysAsp	GlyLeuTrpAlaValLeuAlaTrpLeuSer	IleLeuAlaThrArgLysGlnSerValGlu	ATCTAGCCACCAGCAGAGTGTGGAG	1292
	CTGTCCCTTTGGGAGGAGAGCTTCGGG	ACCGTTCCTGACCACTCCGTTGAGAAAGAT	GGACTGTGGGCTGCTCTGCTGCTCTCC			
	C	T	G	G	T G	A
	AspIleLeuLysAspHisTrpGlnLysHis	GlyArgAsnPhePheThrArgTyrAspTyr	GluGluValGluAlaGluGlyAlaAsnLys	MetMetLysAspLeuGluAlaLeuMetPhe	ATGATGAAGCATTGGAGGCCCTGATTTT	1412
	GACATTCCTCAAAGATCTGGCAAAGCAT	GGCCGGAATTTCTTCAACCGGTATGATTT	GAGGAGGTGGAAGCTGAGGGCGCAAACAA			
	C	C C C	C C	C C	C C	C C
	AspArgSerPheValGlyLysGlnPheSer	AlaAsnAspLysValTyrThrValGluLys	AlaAspAsnPheGluTyrSerAspProVal	AspGlySerIleSerArgAsnGlnGlyLeu	GATGGAAGCATTTCAGAAATCAGGGCTTG	1532
	GATCGCTCCCTTTGGGGAGCAGTTCTCA	GCAAATGACAAAGTTTACACTGTGGAGAAG	GCCGATAACTTTGAATACAGCCACCCAGTG			
	C	G G G	A T C	G C A T	G	G C C A C
	ArgLeuIlePheThrAspGlySerArgIle	ValPheArgLeuSerGlyThrGlySerAla	GlyAlaThrIleArgLeuTyrIleAspSer	TyrGluLysAspValAlaLysIleAsnGln	TATGAGAAGACGTTGCCAAGATTAAACAG	1652
	CGCTTCATTTTCAAGATGGTTCGAAATC	GTCTTCCGACTGAGCGCACTGGAGTGGCC	GGGGCCACCATTCGGCTGTACATCGATAGC			
	G	TG	C A	T C	G	A A
	AspProGlnValMetLeuAlaProLeuIle	SerIleAlaLeuLysValSerGlnLeuGln	GluArgThrGlyArgThrAlaProThrVal	IleThr	ATCACCTaagaagcagccgatgtgtgta	1772
	GACCCCAAGTTCATGTTGGCCCGCTTATT	TCCATGTCTCTGAAAGTGTCCAGCTGCAG	GAGAGAGCGGAGCGCACTGACCCACTGTC		taagaacacagaccagatgatgta	
		T	G A A C A	A A	C	
	cgctccctccaccccgaccatccaagtc	atctgattgaagcagatgacagaacaaaa	tgtattccaagcattttaggatttgact	ttttactaaccagttgacgagcagtgcat		1892
	cgctccctccgcccccaagtcactctgattgaa	gagcatggaagaaacaaagtggataggcc	ccgactttctgggatttgatctttacacta	actgttgccaacacgycgcttctgaggca		
	ttacaaggcactgccccaaagatgcccctt	gggagctgtgagggaagaggacctcgggg	cttagatcaatctcaattctttcatgcc	ctctcgactgtgctgctgctgggtatttgt		2012
	ctgctggttgagagcccttgggagccacg	tgggaaggaggagggggccccgaggtttct	gctaacgtcagtcctctgatcctgccccct	gcactgccccgggtgggtgcttgcctcct		
	ctccttagccatcaggtacagttctacta	caatgtaagctataggtggagcatcagcag	tgagtggagccattcttcatccttaggatg	tggcaatgaaatgagtggtgcaagttcctt		2132
	agcctcaggtacactcagctgtctgtgaa	gyaggcaggtggcattccagcagagggcccc	tttccactctggggctatgcccagtgcaagtg	ccagtcagctgtcttcccccttggaaatc		
	ctctttgtgaaatcttccccattcttct	gtttacatgtaacccacaaatgcaattt	ctagtgcctctctgccaatcagttctcttc	tctgagtgagcgtacttggtactagattt		2252
	gttccccattactgtttacattttacc	aacaaagtgaacttctggtgctctctctc	caatcagttttctcagagtgagacgtac	tcgtctagatttctgcttgttttgaaca		
	ctgccttgttttgacattgtccattca	cacagatattttgggataataaagggaanaa	aaagctaca	2320		
	tggtccgctacacaatttttgggctat	taaaagcaaaataagttgcta				

FIG. 1. Complete nucleotide sequence of the human (Hu) and rabbit (Ra) cDNAs and the derived amino acid sequence for human PGM1. All nucleotides are given for the untranslated regions of both species. The rabbit sequence is shown only at the positions where the amino acid coding sequences differ. Polyadenylation signals and the 25-mer oligonucleotide primers used for PCR amplification are underlined.

The cDNA clone λHPGM1.1 was used as probe to isolate a number of genomic recombinants from Lorist B and λ2001 libraries (Y.H.E. and D.B.W., unpublished data). One of the cosmid clones LoHPGM1 was biotinylated and used as probe

to determine the chromosome location by *in situ* hybridization. After hybridization the banded spreads were aligned with the fluorescent signal (Fig. 4 Lower) and 10 metaphase spreads were scored. The distances from the terminus to the

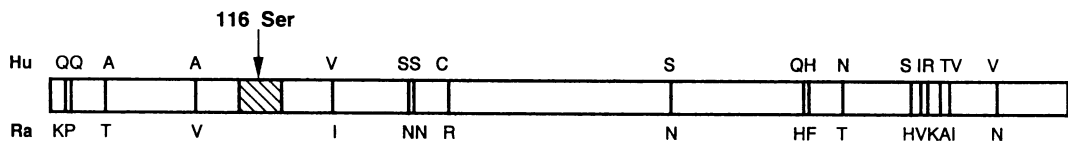


FIG. 2. Diagram of the PGM1 protein showing the positions and identity of the 18 amino acid differences between human (Hu) and rabbit (Ra) PGM1, indicated by the single-letter code and the active site (ref. 23, hatched box in figure) with Ser-116.

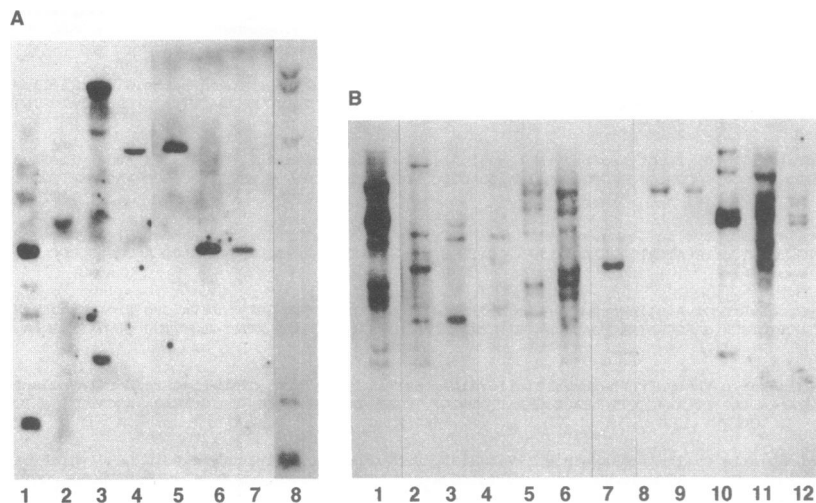


FIG. 3. Southern blot analysis of 15 μ g of genomic DNA from various vertebrate species digested with *Hind*III (A) or *Taq* I (B) and hybridized with 32 P-labeled human PGM1 cDNA. The human *Hind*III digest was of 5 μ g of DNA. Samples in A are as follows. Lanes: 1, zebra; 2, elephant; 3, swan; 4, tree frog; 5, bat; 6, orangutan; 7, common marmoset; 8, human. Samples in B are as follows. Lanes: 1, human; 2, dog; 3, cat; 4, cow; 5, sheep; 6, goat; 7, pig; 8, hamster; 9, rat; 10, mouse; 11, chicken; 12, *Xenopus*.

signal and the terminus to the centromere were measured and expressed as a proportion in each case and gave a value of 0.52 ($n = 20$, $SD = 0.04$). This indicated a location for the PGM1 gene of 1p31.

DISCUSSION

The PGM1 protein is highly polymorphic in humans and an investigation of genetic variation at the nucleotide level should provide detailed information on the evolution of this variation. We are particularly interested to test the hypothesis of Carter *et al.* (11) developed in detail by Takahashi *et*

al. (13), which proposed that reciprocal intragenic recombination between existing amino acid substitutions has been a major mechanism in the generation of structural allelic variants of PGM1 detected by isoelectric focusing. The isolation of the complete human PGM1 cDNA sequence enables the analysis of haplotypes in genomic DNA and the identification of candidate sites of recombination. Our preliminary analysis suggests the total size of the PGM1 gene does not exceed 34 kilobases and thus any tendency for high recombination frequency would seem to be a specific property of the PGM1 sequence rather than a result of large target size.

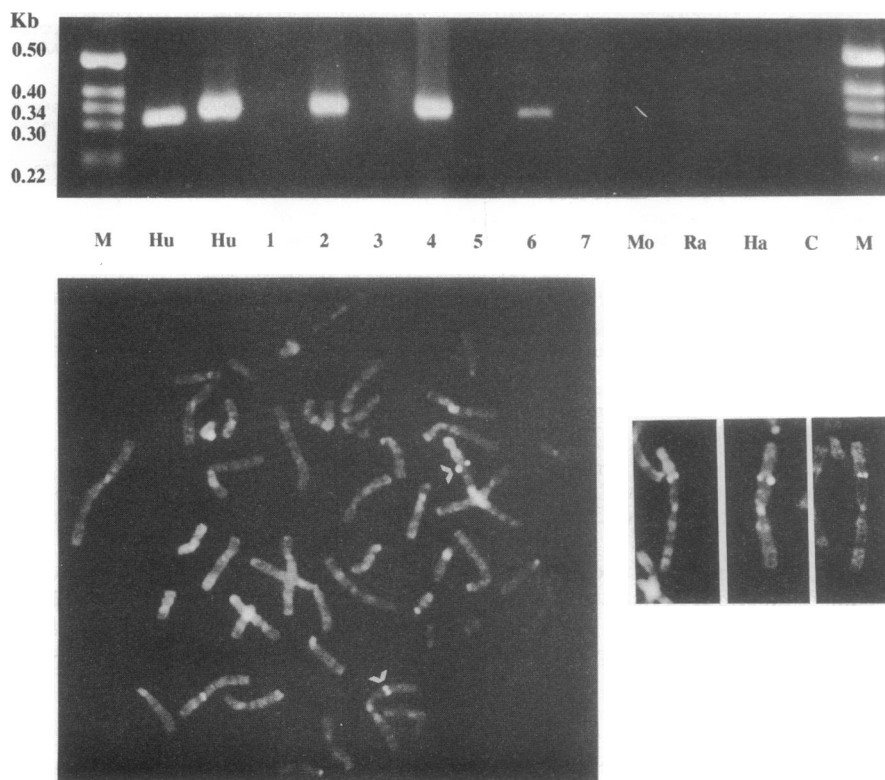


FIG. 4. (Upper) PCR using oligonucleotides specific for human PGM1 and DNA from human (Hu), hamster (Ha), mouse (Mo), rat (Ra), and various rodent-human hybrid cell lines as follows. Lanes: 1, TWIN19F6; 2, TWIN19F9; 3, HCH5; 4, HCH7; 5, CRAB4; 6, CRAB5; 7, CRAB11. A control, no genomic DNA, sample (C) and size markers (M) are also shown. (Lower) Human metaphase R-banded chromosomes showing fluorescent *in situ* hybridization using biotinylated LoHPGM1.1 as probe. In the metaphase spread arrows indicate chromosome 1.

Interestingly, interspecific comparisons indicate less variability than might be expected if this ancient gene were highly mutable. The homology between human and rabbit PGM1 is 92% at the nucleotide level and 97% at the derived amino acid level; furthermore, a high level of sequence conservation appears to be retained throughout the vertebrate kingdom. It is noteworthy that the signal intensity from Southern blot analysis of tree-frog DNA equaled that of the orangutan and marmoset after hybridization at moderate stringency. Such sequence conservation is compatible with the view that the human PGM1 polymorphism has arisen by mechanisms other than the straightforward accumulation of point mutations (11, 13).

A computer-based comparison of the human PGM1 amino acid sequence with those of the human muscle-specific phosphoglycerate mutase PGAM2 (29) and diphosphoglycerate mutase DPGAM (30) revealed no homologies. This is in line with the finding of Ray *et al.* (23) that there are no obvious similarities between the protein sequences of rabbit muscle PGM and yeast phosphoglycerate mutase (PGAM).

No evidence for PGM1-related sequences was found either by Southern blot analysis of human genomic DNA or by *in situ* hybridization. Thus if the genes encoding human PGM2 and PGM3, assigned to chromosomes 4 and 6, respectively (2), arose by duplication from the same ancestral gene as PGM1, then it seems likely that less than 65% sequence homology has been preserved. Immunological data support this view; the anti-rabbit PGM1 antibodies recognize all of the human PGM1 isozymes but not those encoded by PGM2 and PGM3 (31) and no cross reactivity between human PGM1 and PGM2 using anti-human reagents has been demonstrated (32). Thus it may not be possible to identify cDNAs for PGM2 and PGM3 using PGM1 cDNA as a probe, but it is hoped that an oligonucleotide encoding the active site region will be a useful tool for this purpose.

Cloning the human PGM1 gene has allowed the direct chromosomal localization of this locus and this has resolved some uncertainty about the exact position of this widely used genetic marker. A position for PGM1 of 1p22 is generally accepted, although a more distal location has not been excluded (28). In the recent CEPH consortium linkage map of human chromosome 1, assembled by multipoint analysis of 58 loci, PGM1 is positioned distal to the markers D1S22 and ACADM and proximal to D1S19 (33). Pakstis *et al.* (34) using linkage analysis in seven non-CEPH pedigrees, reported essentially the same order but argued that since the SRO (shortest region of overlap) for D1S19 is 1pter-p31 and for ACADM is 1p31 then PGM1 cannot be in 1p22 but rather must be in 1p31 (or more distal). Our results support this view and indicate a localization of 1p31.

We thank Dr. S. Povey and Dr. B. Carritt of the Medical Research Council Human Biochemical Genetics Unit for DNA from somatic cell hybrid lines, Dr. H. Stanley of the Institute of Zoology and Dr. N. Spurr of the Imperial Cancer Research Fund for animal DNA samples, and the Fetal Tissue Bank of the Royal Marsden Hospital for fetal skeletal muscle. Dr. J. Wolfe and B. Cachon of University College London kindly provided access to the Lorist B genomic library. The confocal laser microscope was purchased with a United Kingdom Human Genome Mapping Project grant. M.H. is supported by a grant from the Home Office.

1. Parrington, J. M., Cruickshank, G., Hopkinson, D. A., Robson, E. B. & Harris, H. (1968) *Ann. Hum. Genet.* **32**, 27-34.
2. McAlpine, P. J., Stranc, L. C., Boucheix, C. & Shows, T. B. (1990) *Cytogenet. Cell Genet.* **55**, 5-76.
3. Cantu, J. M. & Ibarra, B. (1982) *Science* **216**, 639-640.
4. McAlpine, P. J., Hopkinson, D. A. & Harris, H. (1970) *Ann. Hum. Genet.* **34**, 177-185.
5. McAlpine, P. J., Hopkinson, D. A. & Harris, H. (1970) *Ann. Hum. Genet.* **34**, 61-71.
6. Quick, C. B., Fisher, R. A. & Harris, H. (1974) *Eur. J. Biochem.* **42**, 511-517.
7. McAlpine, P. J., Hopkinson, D. A. & Harris, H. (1970) *Ann. Hum. Genet.* **34**, 169-175.
8. Fisher, R. A. & Harris, H. (1972) *Ann. Hum. Genet.* **36**, 69-77.
9. Spencer, N., Hopkinson, D. A. & Harris, H. (1964) *Nature (London)* **204**, 742-745.
10. Roychoudhury, A. K. & Nei, M. (1988) *Human Polymorphic Genes World Distribution* (Oxford Univ. Press, New York), pp. 105-110.
11. Carter, N. D., West, C. M., Emes, E., Parkin, B. & Marshall, W. H. (1979) *Ann. Hum. Biol.* **6**, 221-230.
12. Sensabaugh, G. F. & Crim, D. (1982) in *Forensic Science Handbook*, ed. Saferstein, R. (Prentice-Hall, New Jersey), pp. 338-415.
13. Takahashi, N., Neel, J. V., Satoh, C., Nishizaki, J. & Masunari, N. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6636-6640.
14. Whitehouse, D. B., Drago, G. A., Hopkinson, D. A., Westwood, S. A. & Werrett, D. J. (1989) *Forensic Sci. Int.* **41**, 25-34.
15. Lloyd, J. C., Isenberg, H., Hopkinson, D. A. & Edwards, Y. H. (1985) *Ann. Hum. Genet.* **49**, 241-251.
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
17. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266-267.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
19. Henikoff, S. (1984) *Gene* **28**, 351-359.
20. Lichter, P., Tang, C.-J. C., Call, K., Hermanson, G., Evans, G. A., Housman, D. & Ward, D. C. (1990) *Science* **247**, 64-69.
21. Carritt, B., King, J. & Welch, H. M. (1982) *Ann. Hum. Genet.* **46**, 329-335.
22. Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A. (1987) *Ann. Hum. Genet.* **51**, 269-288.
23. Ray, W. J., Jr., Hermodson, M. A., Jr., Puvathingal, J. M. & Mahoney, W. C. (1983) *J. Biol. Chem.* **258**, 9166-9174.
24. Ray, W. J., Jr., & Peck, E. J., Jr. (1972) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), 3rd Ed., Vol. 6, pp. 407-477.
25. Arredondo-Vega, F. X., Charlton, J. A., Edwards, Y. H., Hopkinson, D. A. & Whitehouse, D. B. (1989) *Ann. Hum. Genet.* **53**, 157-167.
26. Gillespie, G., Lloyd, J., Hopkinson, D. & Edwards, Y. (1984) *J. Muscle Res. Cell Motil.* **5**, 457-464.
27. Harris, H. (1980) *The Principles of Human Biochemical Genetics* (Elsevier/North-Holland Biomedical, Amsterdam), p. 78.
28. Bruns, G. A. P. & Sherman, S. L. (1989) *Cytogenet. Cell Genet.* **51**, 67-90.
29. Shanske, S., Sakoda, S., Hermodson, M. A., DiMauro, S. & Schon, E. A. (1987) *J. Biol. Chem.* **262**, 14612-14617.
30. Joulin, V., Peduzzi, J., Romeo, P.-H., Rosa, R., Valentin, C., Dubart, A., Lapeyre, B., Blouquit, Y., Garel, M.-C., Goossens, M., Rosa, J. & Cohen-Şolal, M. (1986) *EMBO J.* **5**, 2275-2283.
31. Drago, G. A., Hopkinson, D. A., Westwood, S. A. & Whitehouse, D. B. (1991) *Ann. Hum. Genet.* **55**, 263-271.
32. Piatti, E., Accorsi, A., Piacentini, M. P. & Fazi, A. (1990) *Biochem. Int.* **21**, 155-163.
33. Dracopoli, N. C., O'Connell, P., Elsner, T. I., Lalouel, J. M. & White, R. L. (1991) *Genomics* **9**, 686-700.
34. Pakstis, A. J., Kidd, J. R., Sparkes, R. S. & Kidd, K. K. (1989) *Cytogenet. Cell Genet.* **51**, 1056 (abstr.).