THE ISOLATION OF THE HUMAN 6-PHOSPHOGLUCONATE DEHYDROGENASE GENE, A GENETIC MARKER IN MAN

by

Patrick Kleyn

A thesis submitted for the degree of
Doctor of Philosophy in the
University of London

January 1990

Department of Genetics and Biometry
The Galton Laboratory
University College London

ABSTRACT

6-Phosphogluconate dehydgogenase (PGD) is an NADP-dependent dehydrogenase involved in the metabolism of glucose-6-phosphate via the pentose pathway.

The enzyme is a dimer which displays an electrophoretic polymorphism which has been widely used in genetic mapping and which allowed its initial assignment to chromosome 1p.

This thesis describes the isolation of human PGD cDNAs. Their identity was established by a combination of genetic and physical mapping studies which indicated their close genetic linkage to the PGD protein polymorphism in the p36 region of chromosome 1.

Sequence analysis of the human cDNAs revealed extensive homology with the *E.Coli* 6-phosphogluconate dehydrogenase gene at the nucleotide level and with the sheep PGD protein at the amino acid level. However, the latter homology required some rearrangement of the cyanogen bromide peptides in the published sequence.

The presence of at least two PGD-related loci in humans is described which are probably processed pseudogenes; one of these was assigned to chromosome 18q.

The human PGD cDNA was shown to detect a restriction fragment length polymorphism (RFLP) in BamHI-cleaved human DNA. This RFLP yielded a polymorphism information content (PIC) of 0.30 compared to 0.04 for the protein polymorphism, a useful increase in

genetic variation detectable at this locus.

The genomic organisation of the human PGD gene was established in outline and a region enriched in hypomethylated CpG dimers was identified at the 5' end.

Physical linkage to the ENO1 locus, which is genetically linked to PGD at a distance of 1.5cM was investigated by pulsed-field gel electrophoresis. Preliminary evidence was obtained suggesting that ENO1 and PGD are separated by less than 350kb, providing a provisional estimate for the relationship between genetic and physical distance in the distal region of the short arm of chromosome 1.

LIST OF FIGURES PAGE

LIST OF TABLES

To my parents,

whose support and encouragement

made this possible

ACKNOWLEDGEMENTS

I should like to thank my supervisor, Dr. B. Carritt for his constant guidance and support throughout the course of this project. I am also grateful to Dr. J.D.A. Delhanty for all her help.

Thanks are extended to K. Gulati and D. Griffin for carrying out the *in situ* hybridisation experiments, to Dr. G. Corney for collection of blood samples and Prof. T. Jenkins for providing blood samples from South African families. I am also indebted to S. Xenophontos for all her assistance.

I thank Dr.D.A. Hopkinson and the MRC Human Biochemical Genetics Unit for providing me with the opportunity to do this PhD.

I acknowledge the Cancer Research Campaign which provided the financial support for the research contained in this thesis.

I wish to thank George Conyne for his moral support and last, but most of all, Julie Blight for her sympathy and encouragement when needed.

ABBREVIATIONS

BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
DEP	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
EDTA	Ethylene diaminetetraacetic acid
G6PD	Glucose-6-phosphate dehydrogenase
GSH	Glutathione (reduced form)
GSSG	Glutathione (oxidized form)
IPTG	Isopropyl-β-D-thiogalactopyranoside
MEM	Minimum Essential Medium
MTT	3-[4,5-Dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide
NAD^+	Nicotinamide adenine dinucleotide (oxidized form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate (oxidized form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
PAM	Accepted point mutations per 100 residues
PEG	Polyethylene glycol
PFGE	Pulsed-field gel electrophoresis
PGD	6-Phosphogluconate dehydrogenase
PMS	Phenazine methosulphate
PUFA	Polyunsaturated fatty acid
PVP360	Polyvinyl pyrrolidone 360
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
T_3	Triiodothyronine
T_4	Tetraiodothyronine (thyroxine)
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
Xgal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CHAPTER 1 INTRODUCTION

1.1 The Pentose Pathway.

6-phosphogluconate dehydrogenase (PGD) catalyzes the oxidative decarboxylation of 6-phosphogluconate (6-PG) to ribulose-5-P. This is the final reaction of the oxidative portion of the pentose pathway which is also referred to by numerous other names including the pentose phosphate pathway, the hexose monophosphate shunt, the Warburg-Lippman-Dickens pathway and the phosphogluconate pathway. I shall refer to it simply as the pentose pathway.

The two major functions of the pathway in man are firstly, the generation of NADPH as a source of reducing power for reductive biosynthesis of fatty acids and steroids, and secondly, the formation of pentoses from hexoses as precursors for nucleotide and nucleic acid synthesis.

This section serves as a brief summary of the reactions of the pentose pathway, the contribution of the pathway to carbohydrate metabolism and its regulation.

Pasteur discovered that yeast is able to break down carbohydrate anaerobically to ethanol or aerobically to CO_2 and water. Furthermore, he showed that the presence of oxygen inhibited the formation of ethanol (Pasteur Effect). When the pathway of glycolysis began to be

investigated, the question arose as to whether the two means of carbohydrate catabolism in yeast were independent or related (Lippman,1936; Dickens,1936). Warburg and his co-workers discovered the enzymatic oxidation of glucose-6-phosphate (G6P) and 6-phosphogluconate by yeast extracts and red cells (Warburg et al.1935, quoted in Lippmann,1936). They showed that both reactions required the presence of a heat-stable co-factor distinct from, but related to the NAD co-factor (the cofactor associated with glycolysis). It was later shown to be NADP (previously called TPN). These two dehydrogenase activities were the first discovered to require NADP as an electron acceptor. The initial interpretation of these findings was that the oxidation of G6P was the first step of aerobic catabolism which was independent of glycolysis. Thus, NAD and NADP were seen as appropriate markers for anaerobic and aerobic metabolism respectively.

Further confirmation of this pathway was provided by Dickens (1936) who showed that yeast would readily metabolize ribose-5-P. This seemed an obvious product of the PGD reaction. In fact, Horecker and co-workers later showed the initial product to be the keto form, ribulose-5-P which is susequently converted to ribose-5-P (Horecker & Smyrniortis,1951; Horecker et al.1951). The discovery of a lactone intermediate (Cori & Lippmann,1952) and the enzyme lactonase which catalyzes its hydrolysis to 6-PG (Brodie & Lippmann,1955)

completed the sequence of reactions of the oxidative part of the pentose pathway as shown in figure (i):

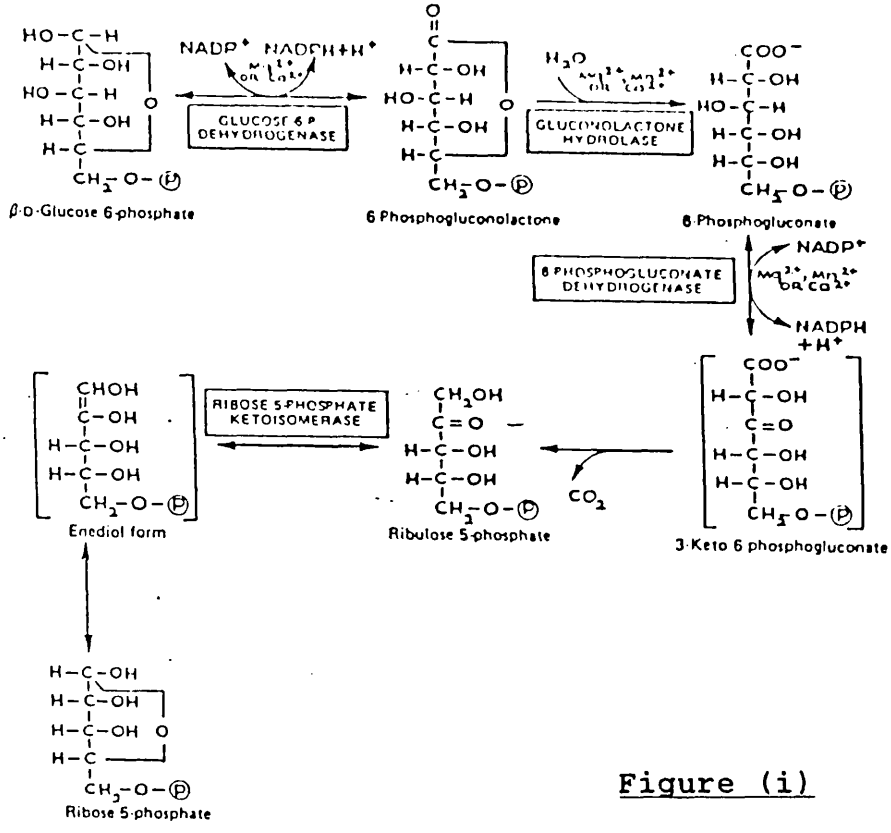


Figure (i)

The oxidative decarboxylation of 6-PG by PGD is presumed to proceed via a 3-keto intermediate as illustrated in fig.1, but this has never been isolated.

By this time, the glycolytic sequence (EMP) had largely been elucidated, Krebs had postulated the TCA cycle and there was growing evidence that these were coupled to oxidative phosphorylation in mitochondria (Axelrod,1967 and references therein). The extensive use of yeast and muscle tissue (both have low levels of pentose pathway activity) for the study of metabolism was in part responsible for the prevailing view that the Embden-Meyerhof pathway was the major or only pathway in most organisms. Thus there was little interest in

alternative pathways. It was not until 1951, when Lehninger showed that the coenzyme of oxidative phosphorylation was NAD and not NADP that it became obvious that the pentose pathway is not the major route of carbohydrate oxidation. By this time several discoveries had rekindled interest in the pentose pathway:

1. Avery, MacLeod and McCarty (1944) had demonstrated that nucleic acid was capable of carrying genetic information. This renewed interest in pentoses which were known to be precursors in nucleotide and nucleic acid biosynthesis.

2. Tracer studies had demonstrated the importance of alternate pathways in *E.coli* and the absence of the EMP had been demonstrated in *L.mesenteroides* (DeMoss et al.1951 quoted in Wood,1955).

Rapid progress in the early fifties mainly in the laboratories of Horecker, Dickens and Racker led to the construction of a scheme for the latter, non-oxidative part of the pentose pathway (see Horecker,1988 and Williams,1987 for reviews). This is illustrated in figure (ii).

The ribulose-5-P produced by the PGD reaction is isomerized to xylulose-5-P and ribose-5-P . Transketolase catalyzes the transfer of a dihydroxyacetone group from the former to the latter to yield sedoheptulose-7-P and glyceraldehyde-3-P. Transaldolase acts on these products to produce fructose-6-P and erythrose-4-P. Transketolase can also transfer a dihydroxyacetone group

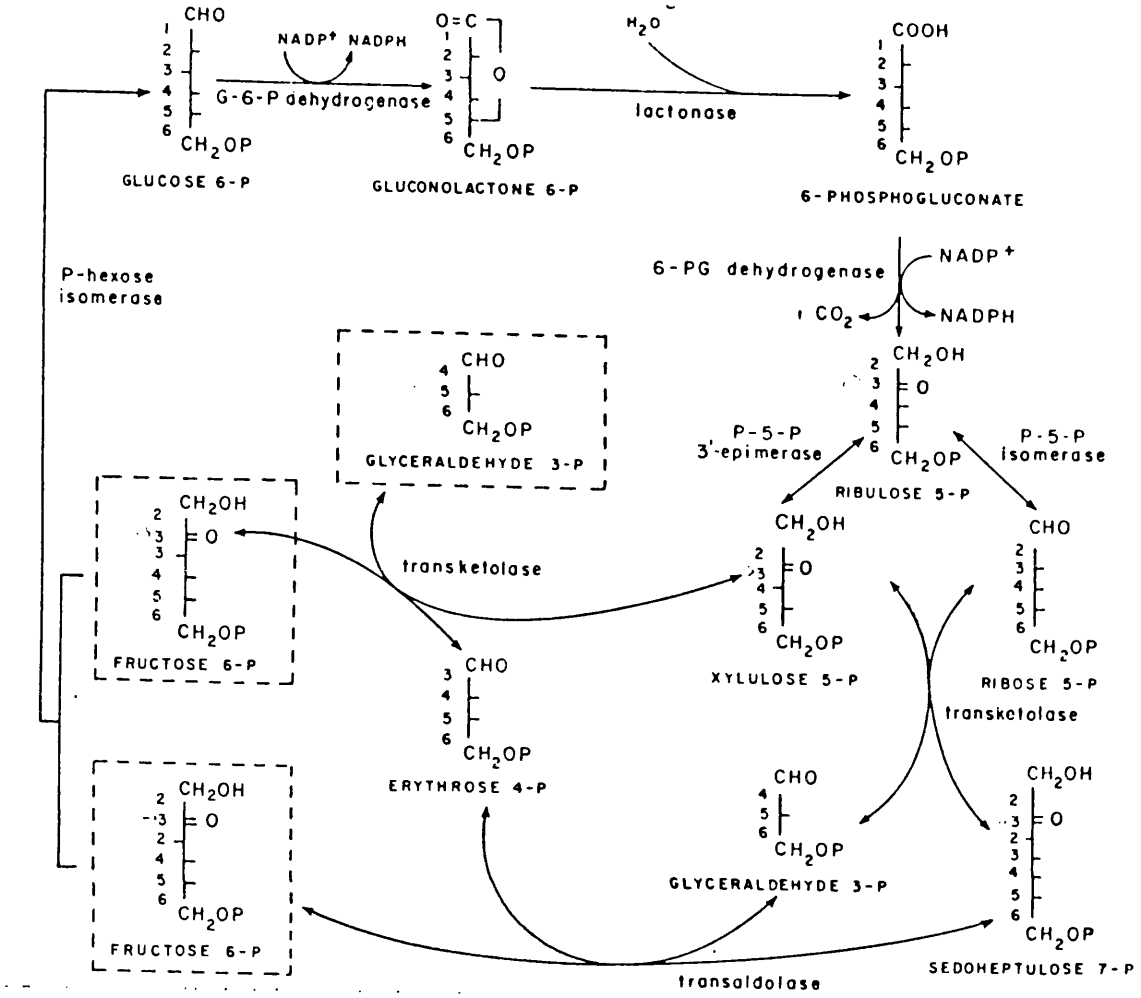


Figure (ii). The 'Classical' or F-type Pathway

This figure shows the reaction sequence for the pentose cycle which, until Williams (1980) proposed an alternative scheme, was universally accepted. The final product of the oxidative part of the pathway is ribulose-5-P. The non-oxidative part involves isomerisation of this into other pentoses followed by various rearrangements which result in the regeneration of phospho-hexoses.

Phosphohexose isomerase allows for at least the theoretical possibility of these re-entering the oxidative pentose pathway. Re-cycling is illustrated on the left.

The numbering refers to the order of the carbon atoms in the original glucose-6-phosphate molecule. It illustrates the rearrangement of this order that is particular to the pentose pathway. Specifically, this scheme predicts that C-2 will become distributed between positions 1 and 3 in the final fructose-6-P molecules in a ratio of 2 to 1 (assuming no recycling).

This diagram is reproduced from Landau & Wood (1983). The reaction products are enclosed in dashed lines.

from xylulose-5-P to erythrose-4-P producing fructose-6-P and glyceraldehyde-3-P. These two products are capable of entering glycolysis which also occurs in the extra-mitochondrial cytoplasm.

Tracer experiments in which [1-^{14}C]-glucose-6-P is used as a substrate will result in the formation of $^{14}CO_2$ (Katz & Rognstad,1966) showing that C-1 is the carbon lost during the decarboxylation of 6-PG. In glycolysis, the fate of the C-1 and C-6 is identical - both become the C-3 of glyceraldehyde-3-P. This difference between the two pathways has been used as a means of assaying the presence of the pentose pathway: Tissues are given a substrate of either [1-^{14}C] or [6-^{14}C]-glucose and the amount of $^{14}CO_2$ measured in each case. A ratio greater than one for $^{14}CO_2$ [1-^{14}C]:$^{14}CO_2$ [6-^{14}C] is indicative of the presence of the pentose pathway.

The isomerisation of fructose-6-P to glucose-6-P by phosphohexose isomerase allows for the cyclical functioning of the pathway (Wood et al.1963) which can be summarized as:

3 Glucose-6-P -------> 3 CO_2 + 3 Pentose-P

3 Pentose-P -------> 2 Fructose-6-P + Glyceraldehyde-P

2 Fructose-P -------> 2 Glucose-6-P

NET: 1 Glucose-6-P --------> 3 CO_2 + Glyceraldehyde-P

Horecker & Mehler (1955), on the other hand, viewed the pentose pathway as two parallel means of providing

pentoses for nucleotide synthesis:

1) by oxidative decarboxylation of 6-PG.

2) by reversal of the transaldolase-transketolase reaction.

Evidence in favour of this view comes from labelling experiments in which it is shown that the pentoses of nucleic acids are derived from both routes (see Horecher & Mehler,1955). Other work, however, suggests that the amount of pentoses produced by the oxidative pentose pathway exceeds by far the requirement for nucleotide synthesis (Wood et al.,1963).

The view taken of the pathway is important since it affects the calculation of the contribution of the pathway to carbohydrate metabolism. Most methods of estimating this depend on the particular reshuffling of carbon atoms that only occurs in the pentose pathway. For instance, a frequent method of measuring the activity of the non-oxidative pentose pathway is to use [1-^{14}C]-ribose-5-P (equivalent to using [2-^{14}C]-glucose) as substrate. The scheme in figure(ii)predicts that two thirds of the ^{14}C passing through the pentose pathway will be found in the 1-position of fructose-6-P with the remaining third in the 3-position. If recycling occurs via phosphohexose isomerase and the oxidative pentose pathway, [2-^{14}C]-Ribose-5-P will be formed (via [3-^{14}C]-glucose-6-P) which can again act as substrate and therefore complicate interpretation. The use of acetone-dried tissue preparations (which are free of NADP)

inhibits such recycling (Horecker et al.1954).

The scheme of the pathway presented in figure(ii)was largely elucidated by the mid-fifties and has been universally accepted since then. Recently, however, Williams and his co-workers (Williams,1980; Arora et al.1985) have cast doubt on the existence of this sequence of reactions *in vivo*. Williams proposes that it is restricted to fat tissue and refers to it as the F-type pathway. He further proposes an alternative pathway (L-type) which he claims operates in liver and other tissues (Williams,1980). This pathway is illustrated in figure 3. It should be emphasized that these pathways are identical with respect to their oxidative reactions. The L-type pathway however, involves 8-carbon intermediates and crucially, predicts incorporation of ^{14}C from [1-^{14}C]-ribose-5-P into C-2 of glucose-6-P. The debate as to the existence of these two pathways continues (Williams et al.1987; Landau,1989) and is largely based on the differing interpretations of tracer experiments. Although these are conceptually simple, they prove to be difficult to analyze for the reasons previously mentioned. These experiments will not be further discussed since they have no direct bearing on the reaction catalyzed by PGD. It should be noted, however, that if the L-type pathway does play a significant role *in vivo*, the estimates of the pathway based on the ^{14}C-1/^{14}C-3 ratio experiments would not be valid.

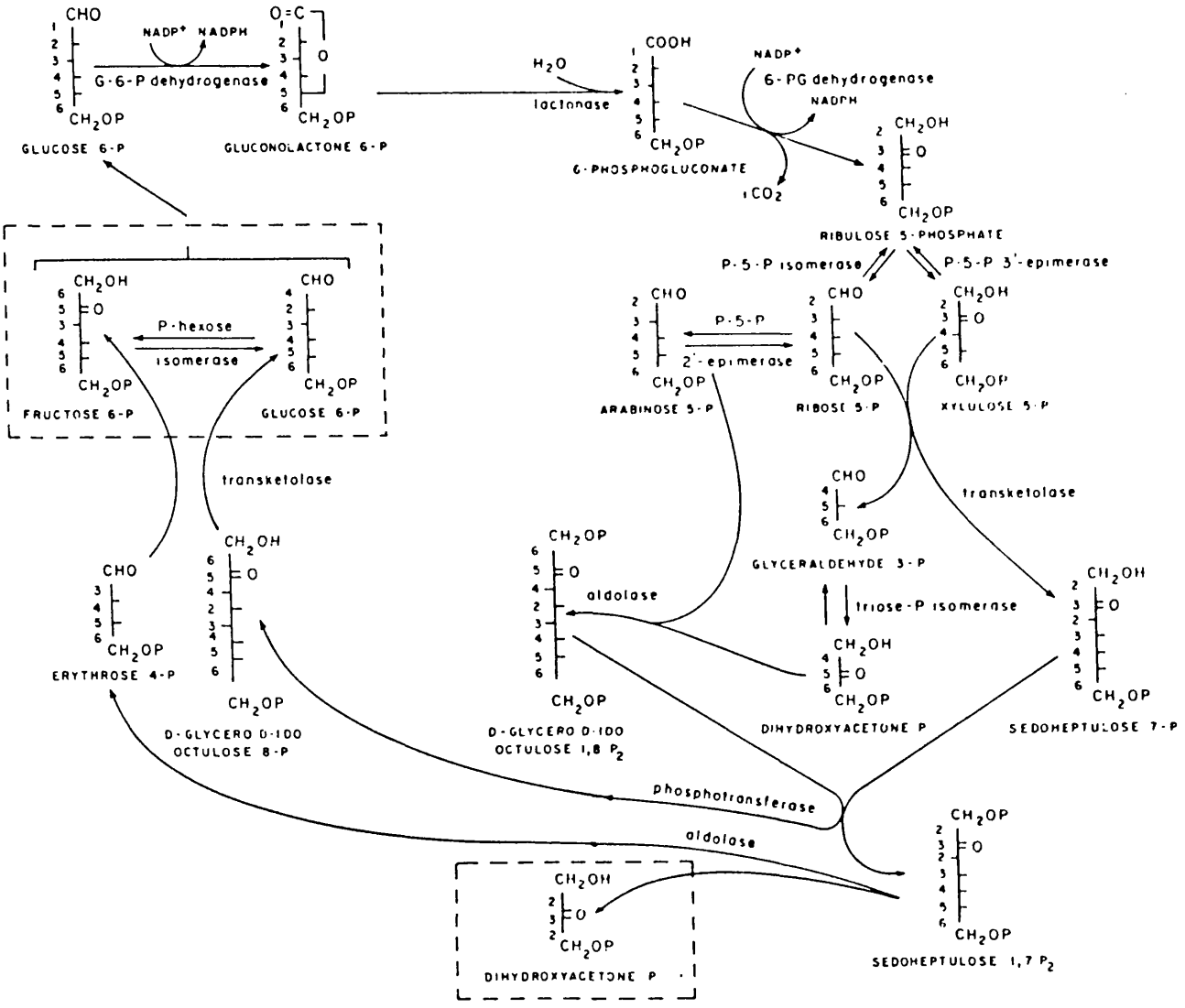


Figure (iii). The L-type Pathway

This figure shows the reaction sequence for the pentose cycle proposed by Williams (1980). It can be seen that the oxidative pentose pathway reactions are identical to those in the F-type or 'Classical' pathway (figure ii). The difference lies in the products formed from ribulose-5-P. This scheme predicts the formation of an 8-carbon intermediate and importantly it predicts that the C-2 carbon of the original glucose-6-phosphate molecule becomes the C-2 of the final G6P product. It is this difference with the F-type scheme which theoretically makes them readily distinguihable.

This diagram is adapted from Landau & Wood,1983. The reaction products are enclosed in dashed lines.

The present consensus is that, in general, the pentose pathway contributes in a minor way to glucose metabolism (Katz & Wals,1972). The higher levels in liver, adrenal glands and adipose tissue (Glock & McLean,1954) reflect its importance in the generation of NADPH for reductive biosynthesis of fatty acids and steroids (Rognstad & Katz,1979).

In red blood cells, the pentose pathway plays an important role in the maintenance of the redox potential (Eaton and Brewer,1974). The red cell is exposed to high oxygen pressures and is therefore liable to oxidant damage. Protection from this is afforded by glutathione and catalase. The oxidation of glutathione from a reduced state (GSH) to GSSG protects hemoglobin from oxidation of its globin moiety by hydrogen peroxide. This protection is dependent on the continuous reduction of GSSG back to GSH by NADPH which is generated by the pentose pathway. Indeed, Jacob and Jandl (1966) suggested that the activity of the pentose pathway in red blood cells is governed by the GSSG/GSH ratio. The maintenance of catalase in an active state capable of reducing H_2O_2 may also be NADPH-dependent (Eaton et al.1972).

The particular importance of the pentose pathway to red cell survival is illustrated by the fact that G6PD deficiency leads to decreased red cell life span and if severe it results in hemolytic anemia. G6PD is most frequently revealed by hemolysis induced by anti-

malarial agents ('primaquine sensitvity'- Carson et al.1956).

The absence of such symptoms in individuals with PGD deficiency suggests that the NADPH generated by the G6PD reaction alone is sufficient for the requirements of the red blood cell and other cells.

1.2 Regulation of the Pentose Pathway.

Since the elucidation of the mechanism of the pentose pathway, interest has turned towards its regulation. The rate-limiting step of the pathway is thought to be glucose-6-phosphate dehydrogenase (G6PD; Kather et al.1972; Kirkman et al.1980; Raivio et al.1981). The level of pentose pathway activity has been found to be strongly influenced by hormonal status and diet composition.

The first demonstration of the influence of diet on the pathway was reported by Glock and McLean (1955a). They showed that rats that have been starved for several days possess reduced level of liver G6PD and PGD activity. Upon refeeding with a high carbohydrate diet, these levels are increased by a factor of 3 to 10 and higher than those of non-starved controls ("overshoot") (Tepperman & Tepperman,1958).

The influence of diet is not simply the effect of changes in concentration of substrate available to the dehydrogenases of the pentose pathway. The use of protein synthesis inhibitors such as puromycin (Potter & Ono,1961) and transcription blockers such as actinomycin D (Connor Johnson & Sassoon,1967) has demonstrated that the induction is dependent on gene transcription and subsequent enzyme synthesis. This has been confirmed by assaying G6PD and PGD mRNA levels using *in vitro* translation and immunoprecipitation (Miksicek &

Towle,1982) or hybridisation to cloned cDNAs (Miksicek & Towle,1983). Thus, carbohydrate induction of the pathway occurs primarily at the pre-translational level although evidence has been presented in support of post-transcriptional regulation in the case of G6PD and other lipogenic enzymes. There is some evidence that, unlike PGD, these are affected by cAMP levels (Sun & Holten,1978; Rudack Garcia & Holten, 1974). However, the more recent consensus is that insulin and glucagon have no effect on the pentose pathway (Proscal et al.1976; Hawthorne & Alberti,1988).

It was shown that the induction of the pathway correlates with that of lipogenesis as assayed by incorporation of acetate-1-^{14}C into lipids (McDonald & Johnson,1965). It had also been discovered that the formation of fatty acids such as palmitic acid from acetyl-CoA required NADPH as a reducing agent. Although the correlation between lipogenesis and the pentose pathway was rapidly established, there still remained the question of the order in which the two pathways were induced. Two obvious possibilities are (Tepperman & Tepperman,1963):

1. The rate of lipogenesis is governed by the availability of NADPH and therefore by the rate of the pentose pathway ("push" hypothesis).

2. An increase in lipogenesis increases the amount of available $NADP^+$ and this leads to increased pentose pathway activity ("pull" hypothesis).

The evidence suggests the latter view to be more appropriate. In particular, the [NADPH/$NADP^+$] ratio seems to be important in determining the rate of pentose pathway activity. For instance, even under conditions of low lipogenesis, the liver pentose pathway can be induced by other NADPH-requiring reactions such as fatty acid chain elongation or desaturation of saturated fatty acids (Tepperman & Tepperman,1965). Also, oxidizing agents such as methylene blue and PMS are capable of strongly inducing the pathway (Kather et al.1972; Raivio et al.1981).

Refeeding of a starved rat with a diet containing unsaturated fat suppresses the induction of the pentose pathway and lipogenesis by carbohydrate (Tepperman & Tepperman,1965; Rudack et al.1971). The components responsible for this effect have been shown to be polyunsaturated fatty acids (PUFA), linoleic or linolenic acid (Nace et al.1979). The metabolism of PUFA suppresses the synthesis of G6PD and PGD mRNAs as well as those encoding other lipogenic enzymes (Tomlinson et al.1988).

Initial measurements of G6PD and PGD activity levels in male and female rats showed them to be consistently higher in females (Glock & McLean, 1953). Glock and McLean interpreted this as an influence of the sex hormones on these dehydrogenases. Huggins and Fao (1959) supported these findings and showed differential responses from PGD and G6PD to estradiol in ovariecto-

mized rats.

It was shown that both PGD and G6PD respond to thyroid hormone and that the response to thyroid hormone requires *de novo* enzyme synthesis (Glock & McLean,1955; Huggins & Fao,1959; Diamant et al.1972;Gibson et al.1972).This was later shown to result from increased levels of G6PD and PGD mRNAs (Miksicek & Towle,1982;1983).

The thyroid hormones, triiodothyronine (T_3) and thyroxine (T_4), and the biosynthetically unrelated, cholesterol-derived steroids seem to act via a very similar mechanism (Weinberger et al.1986). Like the steroids, thyroid hormones penetrate the cell membrane whereupon they bind to receptors which exist in the cytoplasm and the nucleus (Oppenheimer & Samuel,1983). The cellular homologue of the avian oncogene, v-*erbA* (c-*erbA*) encodes a thyroid hormone receptor (Sap et al.1986; Weinberger et al.1986). This gene displays high sequence homology with those for steroid receptors and to an even greater degree with those for vitamin D3 and retinoic acid (see Chambon & Green,1988 for review). This suggests that all these receptors may have evolved from one ancestral gene. Support for this hypothesis comes from analysis of the chromosomal localization of the genes for the thyroid hormone and retinoic acid receptors. Both receptors have at least two genes, each with one on chromosome 3 and one on chromosome 17 (Evans,1988). Deletion analysis and hybrid receptor

experiments have shown that distinguishable domains are responsible for ligand-binding, DNA-binding and transcriptional activation (see Maniatis et al.1987; Evans,1988; Green & Chambon,1988 for reviews). The DNA-binding domain of the thyroid hormone receptor forms a Zinc-finger motif which recognizes a 16bp palindromic sequence which has been found upstream of the growth hormone gene and the G6PD gene. In contrast to the steroid receptors (Evans,1988; Mitchell & Tijan,1989), the thyroid hormone receptor (and v-*erbA*) has been found to bind this sequence in the absence of its ligand. In this case it suppresses the activity of the nearby promoter (Damm et al.1989). Binding of the ligand to the thyroid hormone receptor activates the nearby promoter.

The effects of diet and hormones can be regarded as "coarse" control in which the activities of the pentose pathway dehydrogenases may vary severalfold. These effects are mediated by changes probably primarily at the transcriptional level. As many hepatic, lipogenic enzymes respond in a similar fashion (Diamant et al.1972; Mikscicek & Towle,1982; Hawthorne & Alberti,1988) to fat-free, high carbohydrate diets and thyroid hormones, it seems plausible that these two effectors regulate gene expression via common pathways. However, as yet, there is no evidence of a molecular nature to support this view. Superimposed upon this regulation is the "fine" control at the enzyme level.

It has been shown that the [NADPH]/[$NADP^+$] ratio at the site of G6PD has a critical role in the regulation of the pentose pathway (Kather et al.1972). Ever since the discovery of the pathway it had been known that NADPH had an inhibitory effect on G6PD (Negelhein & Haas,1935, quoted in Eggleston & Krebs,1974). Therefore a high [NADPH]/[$NADP^+$] ratio has a dual inhibitory effect on G6PD: Low levels of $NADP^+$ substrate and high levels of NADPH inhibitor. There were several inconsistencies, however, in the data. It had been found that a [NADPH]/[$NADP^+$] ratio of 9 resulted in nearly 100% inhibition of G6PD *in vitro* and that the ratio in rat liver cytoplasm is approximately 100 (Veech et al.1969). Furthermore, the activity of PGD in rat liver cytoplasm is 5-fold less than that of G6PD under standard conditions and 100-fold less at physiological concentrations of substrates (Sapag-Hagar et al.1973).

These findings suggested two things; firstly, the existence of another factor which interfered with NADPH inhibition of G6PD and PGD. Secondly, there must be a more powerful de-inhibition of PGD relative to G6PD since PGD is not the rate-limiting enzyme of the pathway. Eggleston & Krebs (1974) found that oxidized glutathione (GSSG) had such properties. They found that the suppression of NADPH inhibition by GSSG required the presence of an unstable cofactor. This has been characterized as a protein of M.W. ~ 10^5 which is found in kidney and adipose tissue as well as liver (Garcia et

al.1989). In liver and adipose tissue it was found to be diet-inducible. The cofactor-GSSG complex was found to have a greater disinhibitory effect on PGD than on G6PD. This provides a mechanism for overcoming the greater inhibition of PGD by NADPH. A further level of complexity may be the availability of $NADP^+$ to G6PD. It has been shown that, in human erythrocytes at least, proteins other than the $NADP^+$-linked dehydrogenases bind $NADP^+$ and that the ratio of bound to unbound $NADP^+$ varies with sigmoidal kinetics (Kirkman et al.1986). In conjunction with the finding that intracellular G6PD also displays sigmoidal kinetics and that its activity is under constraint relative to isolated G6PD (Kirkman & Gaetani,1986), this suggests that the pathway is also regulated by these $NADP^+$-binding proteins.

In summary, the coarse control of the pentose pathway is exerted by carbohydrate, PUFA and hormones (primarily the thyroid hormones). This occurs at the pre-translational level. In the case of the thyroid hormones this is mediated by the thyroid hormone receptor. This recognizes specific sequences found within and near to the 5' end of its target genes. Additional to this pre-translational regulation, both G6PD and PGD activities are modulated by the $[NADPH]/[NADP^+]$ ratio which in turn is dependent on the level of NADPH-requiring biosynthesis. There is evidence to suggest that the available $NADP^+$ concentration is also influenced by oxidized glutathione and NADP-binding proteins.

1.3 PGD Enzyme Polymorphism

The PGD enzyme displays a polymorphism which was first reported by Fildes & Parr(1963). The advent of starch gel electrophoresis had already permitted the demonstration of variation in a number of enzymes including glucose-6-phosphate dehydrogenase (Boyer et al. 1962), serum cholinesterase (Harris et al. 1962) and acid phosphatase (Hopkinson et al. 1963).

Fildes and Parr electrophorized hemolysates in 0.01M, pH7.0 phosphate-buffered starch gels. PGD activity was detected by overlaying the gel with 6-phosphogluconate/NADP/PMS/MTT in agar and incubating at 37°C for an hour. The 'normal' or 'A' pattern, in which a single band migrates anodally was seen in most cases. In 10 out of 150 cases, however, a second more slowly migrating band was also observed. They referred to this as the 'B' pattern. Examination of this trait in families suggested that it was controlled by a single, two-allele autosomal locus with simple Mendelian inheritance and that the B pattern represented the heterozgous condition.

The polymorphism can also be revealed in white cell extracts (Parr 1966). In this case and at higher resolution the B pattern shows three bands with relative intensities of 1:2:1. Parr(1966) found in a survey of 4,558 Londoners an incidence of 4% of the B pattern. In two individuals he found a third pattern, C, with only

one slowly migrating band. This suggested that the enzyme is a dimer made up of freely-recombining subunits such that the A and C patterns represent homozygotes for two alleles, PGD^A and PGD^B [1]. The heterozygote, $PGD^A PGD^B$, produces two subunits, S^A and S^B which combine to give three types of dimers, $S^A S^A$, $S^A S^B$ and $S^B S^B$ in the ratio 1:2:1. Thus, heterozygotes give B electrophoretic patterns (Fildes & Parr 1963; Parr 1966; Parr & Fitch 1967). Assuming a Hardy-Weinberg equilibrium and no other alleles, a heterozygote incidence of 4% suggests a minor allele frequency of 0.02.

In the red cell hemolysates in which this polymorphism was first studied, the banding pattern is more complex than it is in white cell extracts. In the $PGD^B PGD^B$ individuals there are 2 minor bands which comigrate with the bands of the heterozygote (Parr & Fitch 1967). The explanation for this is unclear but it may be related to the stromal effect on PGD (Carson. et al. 1966). When red cell hemolysates are incubated with haemoglobin-free red cell membranes (stromata) and NADP, PGD is partially inactivated. Furthermore, electrophoretically detectable change occurs such that the AA pattern resembles the BB pattern after this treatment (Ajmar et al. 1968). This effect was shown to be

---

[1] There has been considerable inconsistency in the PGD allele nomenclature. The minor allele is often referred to as PGD^C. Parr (1966), however, suggested the name PGD^B and this is how it shall be referred to in the following discussion.

mediated by a product of stromal NADPase activity and NADP, namely, 2-phopho-adenosine diphosphate ribose (P-ADPR). The change in the AA PGD dimer, however, results in slower migration of the enzyme. Whether a similar effect could result in faster migration of the BB dimer to produce the effect seen in the electrophoresis of red cells of *PGD*^{B} *PGD*^{B} individuals has not been investigated.

In spite of these effects, the PGD phenotypes are readily scored in hemolysates. The electrophoretic phenotypes of red cell PGD have been assessed in many populations world-wide by a number of groups (Parr 1966; Bowman et al. 1966; Gordon et al. 1967; Shih et al. 1968; Tills et al. 1970 & 1971; Beckman 1972; Jenkins & Nurse 1974; Mourant et al. 1976). Their results are tabulated in Appendix A.

Briefly,the *PGD*^{B} allele frequency, with one exception, ranges from 0.014-0.039 in Caucasian populations giving an average polymorphic information content(PIC)=0.09. Negroid populations range from 0.000 to 0.139. Populations in the East also have higher frequencies: The Middle East, 0.028-0.092; Asiatics, 0.014-0.231 and Negroes, 0.000-0.139. There is a suggestion of a gradient from East to West of decreasing *PGD*^{B} frequencies. Also, both are absent in a number of American Indian populations (Bowman et al. 1966; Tashian et al. 1967). Either this may reflect the mixing of two populations, an Eastern one with *PGD*^{B} and the blood group B gene and a Western one without

both or it may be due to the same local selective pressures on the two genes (Tills et al. 1970).

The difference between *PGD*^{B} frequencies in two populations in Ethiopia living at different altitudes (Harrison et al. 1969) suggested a possible selective advantage of the *PGD*^{B} allele at higher altitude. The high frequency of *PGD*^{B} reported in a population in Bhutan (Carter et al. 1968, Mourant et al.1968, both quoted in Tills et al.1970) is in accordance with this theory. However, the difference found in the study was only significant at the $P<0.05$ level. Furthermore, the high levels of *PGD*^{B} in Swedish Lapps (Beckman 1971) and some African subpopulations (Jenkins & Nurse 1974) do not support the hypothesis.

Soon after the discovery of the *PGD*^{B} variants, many other PGD electrophoretic variants were discovered. Those found to date are summarized in Appendix B. Most of these variants are rare, but it is possible to calculate their gene frequencies in some populations in which they are more frequent. For instance, the Richmond variant (Parr 1966) is rare but widely-distributed. In Parr's London survey of 4,558 individuals two were found to have the Richmond phenotype and it has been found at a frequencies of 0.010 in a native population in Dominica. The low frequency but widespread distribution of *PGD*^{Richmond} suggests either recurrance of the same mutation or recent introduction of the allele by colonisers.

Many of these variants have not been examined side

by side and may be identical. For instance, Richmond and Thai both migrate faster than PGD^A. They may represent identical mutations or may have different alterations which result in the same mobility change (Tuchinda et al. 1968; Blake et al. 1974). Jenkins and Nurse (1974), on the other hand, suggested that Thai may be identical to Elcho. Isoelectric focusing of PGD could resolve such issues and further subdivide some variants, but no such analysis has been carried out to date.

Some of the electrophoretic variants show a change of mobility upon treatment with certain reagents. The Friendship and Natal variants are the most striking examples of this category.

The Friendship variant (PGD^F) can only be detected in the absence of $NADP^+$ in the starch gel (Davidson, 1967). If $NADP^+$ is added to the gel, as is done routinely, the Friendship variant reverts to the A type. It was also found that addition of NADP to the hemolysate or cell extract prior to electrophoresis had no effect. This effect is therefore unrelated to the stromal effect on the Plaistow (AB) and Canning (BB) variants.

The PGD^{Natal} variant produces a pattern similar to $PGD^{Richmond}$, but the former reverts to the A type upon treatment with the thiol reagent, 2-mercaptoethanol (Blake et al. 1974).

Similar changes had previously been observed in other enzyme systems such as in the PepA5-1 variant

(Lewis et al. 1968). This produces a three-banded pattern similar to PepA2-1 when a fresh hemolysate is used. However, after storage (4° C/1 week) faster-migrating bands replace the original bands. This effect can mimicked by the addition of oxidized glutathione (which is known to accumulate in stored red cells Lewis et al.1968) to fresh hemolysates. Furthermore, addition of 2-mercaptoethanol reverts the pattern of stored samples to that of fresh ones. This effect was also investigated with adenosine deaminase (Hopkinson & Harris 1969). These modifiations were interpreted as being mediated by a cysteine substitution in the *Pep*A5-1 and PGD^{Natal} which would introduce a reactive sulphydryl (-SH) group able to react with oxidized glutathione (G-S-S-H):

G-S-S-H + Pr-S-H <=> Pr-S-S-G + G-S-H

The resulting polypeptide has an additional two carboxyl groups and only one additional amino group, thus aquiring a net gain of one unit of negative charge. This model fits with the increased mobility of the subunits upon storage and a reversion of this effect by 2-mercaptoethanol which restores sulphydryl groups by reducing disulphide bonds. This hypothesis was further supported by investigation of the phenomenon in adenosine deaminase (Hopkinson & Harris,1969) and the phosphisomerase variant, PHI5-1 (Hopkinson,1970) with other thiol agents.

AB and BB hemoysates show slightly lower activity than AA: 80-100% and 70-90% respectively (Parr 1966).

The three phenotypes react differently when stressed (Parr & Parr 1965). Incubation of hemolysates of genotype AAin 0.3M Tris.Cl,pH8 at 37°C for 15 min. results in a mean decrease of activity of 27% whereas the equivalent figure for AB is 49% and for BB is 62%. Partial denaturation with urea or treatment with iodoacetate results in a similar differentiation. In both cases the S^B subunits reduce the stability of the dimers. Other evidence, however suggests that the lower stability of the *PGD*^{B} allele *in vitro* does not reflect any physiological effects. Firstly, several deficient and partially deficient PGD variants have been reported and in all but two cases no clinical conditions has been associated with them (Parr & Fitch 1964 & 1967;Dern et al. 1966). This contrasts with G6PD deficiency which is associated with haemolytic anaemia (Carson et al.1956). Secondly, *In vitro* assays which measure the rate of reduction of NADP such as the methemoglobin reduction test (Brewer & Tarlov,1962) that are used to screen for G6PD deficiency give normal results with PGD partially-deficient individuals (Brewer et al.1962). The implication is that if deficiency for activity has no detectable effect, reduced activity because of instability of the *PGD*^{B} allele will be of little consequence. In the individuals with haemolytic anaemia and PGD deficiency referred to above other factors may have been responsible for anaemia. One had raised levels of G6PD indicating a young red cell population (Scialom et

al. 1965, quoted in Eaton & Brewer,1974). The second case, a newborn infant also had decreased GSH, NADH and ATP levels implying a higher order defect affecting the anaerobic pathways as well as the pentose phosphate shunt (Lausecker et al. 1965, quoted in Eaton & Brewer,1974).

The PGD alleles with variant activities form the final category of PGD enzyme polymorphism (Parr,1966; Parr & Fitch,1967).

Parr & Fitch (1964) reported an English family which could be divided into two distinct groups according to PGD activity (as a proportion of both mass of haemoglobin and G6PD activity). Their assay was based on the spectrophotometric measurement of increase in NADPH concentration during the dehydrogenation of 6-phosphogluconate. The partially deficient individuals had activity levels of 50-60% of those of normal relatives and controls. The inheritance suggested that the partially deficient condition (Ilford variant) is a heterozygous one. The electrophoretic pattern is indistinguishable from the A phenotype apart from the reduced intensity of the band. They also found one deficient individual with a Canning (BB) band pattern of reduced intensity which they called Newnham. In this case PGD had slightly reduced activity relative to Ilford variants. The fact that Ilford and 'Normal' both possess single bands of identical mobility and that the same is true of Newnham and Canning implies that the gene responsible for the deficiency does not contribute any subunits to the

dimers (Parr & Fitch 1967). They therefore proposed the genotypes *PGD*^{A} *PGD*^{0} and *PGD*^{B} *PGD*^{0} for the Ilford and Newnham variants respectively.

The possibility that the deficiency was controlled by a separate regulator locus and that Ilford and Newnham were homozygous for electrophoretic alleles at the PGD locus was precluded by the discovery of two *PGD*^{B} *PGD*^{0} - *PGD*^{A} *PGD*^{A} matings in one family. Of the 7 progeny, 5 showed only a single A band of reduced intensity (A0) and the remaining two were common variants (AB). Thus genes encoding the electrophoretic variants and this deficiency were shown to be allelic.

In another family a more severe deficiency allele, *PGD*^{W} was found. In the red cells of homozygotes (Whitechapel variants), the enzyme activity is less than 5% of normal whist in the *PGD*^{A} *PGD*^{W} heterozygotes (Dalston) it is about 75%. Unlike the Ilford and Newnham variants, PGD activity in white cells did not correlate with that in red cells. In the white cells of Whitechapel individuals there was appreciable activity which, however, was less stable under stress. Since there is no protein synthesis in mature erythrocytes, this suggests that the *PGD*^{W} allele is not null but encodes a subunit which is unstable. The observation that heterozygotes for this allele still possess 75% PGD activity (rather than 50%) suggests that the S^{A} subunit confers stability on the S^{A}S^{W} dimer.

Other reports of PGD deficiency (Brewer & Dern 1964

and Dern et al. 1966) seem to fall in one of the two categories discussed above. The only exception to this is the case of one partially-deficient individual with an AB electrophoretic phenotype in which both bands were reduced in intensity suggestive of a regulator gene independent of the structural locus. However, since this individual was the only member of the family with clearly reduced PGD activity, this hypothesis could not be tested.

1.4 Genetic and Physical Mapping

The theory of linear arrays of genes along cytologically visible chromosomes as postulated by Morgan (1911) was supported by the construction of the first genetic linkage map of six X-linked loci of *Drosophila* by Sturtevant (1913). These pioneering works together with Muller's seminal paper on crossing over (1916) laid the foundation for the subsequent genetic mapping of the chromosomes of a number of experimental organisms notably *Drosophila melanogaster*.

The construction of genetic maps requires the analysis of recombination between pairs of markers (or between several markers simultaneously). Organisms such as Drosophila and to a lesser extent mouse are well-suited to this for several reasons:

1. Crosses between double heterozygotes and recessive homozygotes from which all progeny are informative are easily arranged.
2. An unlimited number of meioses can be examined quickly because of the short generation time.
3. Phenotypic polymorphisms with simple Mendelian inheritance are readily available for analysis.

In man, none of these conditions apply and it was therefore not until 1937 that linkage between two genetic markers (colour-blindness and haemophilia) in man was reported (Bell & Haldane,1937). This achievement immediately suggested the theoretical poss-

ibility of disease diagnosis by linkage to readily scored traits. For instance, Morton (1955) detected linkage between the Rhesus locus and elliptocytosis. However, the paucity of phenotypic variation available in Man severely restricted the general application of this approach.

The discovery of extensive phenotypically benign variation at the protein level completely changed the prospects of linkage analysis in man. The newly-discovered class of enzyme variants was extensively investigated in the context of gene mapping and linkage to disease loci. However, in spite of the increase in detectable genetic variation this represented, it was insufficient for the construction of useful genetic maps. This is the consequence of studying variation at the protein level which represents only a subset of the variation present in the genome: Coverage is restricted to that portion of the genome encoding genes and furthermore, only those base substitutions which alter amino acids are detected. Restriction fragment length polymorphisms (RFLPs) represent a means of directly studying variation at the DNA level and therefore provide wider coverage of the genome.

The first restriction fragment length polymorphism (RFLP) to be discovered was at the β-globin locus (Kan & Dozy,1978). This RFLP was found to be in linkage disequilibrium with the β^s sickle-cell anaemia allele in a San Francisco black population. The feasibility of

disease diagnosis using RFLPs was carefully assessed by the authors. They advocated accurate evaluation of the recombination frequency between a marker and disease since this determines the confidence in the diagnosis using the marker. Obviously, only tightly-linked markers are suitable for disease diagnosis and the final aim must be the isolation of a gene probe which detects the genetic defect(s) itself.

Jeffreys (1979) estimated the presence of up to $3x10^7$ silent base substitutions potentially detectable by restriction enzymes. This prompted Botstein et al.(1980) to suggest the realisable goal of constructing a 20cM map of the human genome. A strict spacing like this would permit the addition to the map of any new locus, for example one involved in genetic disease. Assuming the total human genetic length to be approximately 33 Morgans (Renwick,1971), this implies a minimum of 150-200 evenly-spaced markers covering the genome. Obviously to attain such coverage of the genome a far greater number of markers need to be mapped. At the last Human Gene Mapping conference (HGM10,1989), 3370 RFLP markers were reported and for most of the genome a spacing of 20cM or less has been achieved (Keats et al.1989).

For some regions of the genome, maps of resolution much greater than 20cM exist (Barker et al., Keats et al.1989). Most of these maps cover regions to which loci involved in genetic diseases had previously been

localized.

The first autosomal linkage detected as a result of a random RFLP search was between Huntingdon's chorea and G8, a polymorphic probe on chromosome 4p (Gusella et al.1983). This was one of twelve RFLPs studied. Detection of linkage using only twelve markers would not generally be expected since this represents a coverage of, at most, 5 to 10 Morgans and possibly much less. This illustrates the need for a genetic map. The even 20cM map advocated by Botstein et al. implies that no new locus would be more then 10cM from a mapped marker. Clearly in those regions of the genome for which maps of 20cM resolution or better exist, failure to include a new locus can be used to construct exclusion maps. In this way genetic maps of groups of markers have been used to exclude disease loci from large, continuous segments of the genome (Wainwright et al.1986; Bradley et al.1989).

These approaches require no knowledge of the nature of the physiological defect causing the disease apart from a means of accurate differential diagnosis. The disease is simply considered as a genetic marker and its segregation in pedigrees is studied with respect to a panel of random RFLPs. However, an absolute prerequisite is an understanding of the mode of transmission of the disease.

The human genetic diseases investigated to date using linkage analysis have generally been single-gene

defects with simple Mendelian inheritance. Many, if not most diseases with a genetic component, however, have more complex modes of transmission caused by variable penetrance, genetic heterogeneity or the involvement of several loci. Examples of such diseases include the inherited forms of schizophrenia (Sherrington et al.1988; St.Clair et al.1989) and affective disorders (Hodgkinson et al.1987). Single-marker analysis as described above is far less efficent when applied to these diseases. One way of avoiding the problem of genetic heterogeneity (but not that of penetrance or polygenic inheritance) is to study single, large pedigrees in which all the affected individuals might be expected to be carry the same defect (Gusella et al.1983; Egeland et al.1987; Kelsoe et al.1989). The validity of this obviously depends on the incidence in the population of all genes producing the affected phenotype. The same considerations apply to genetic diseases for which there is a significant proportion of sporadic cases.

Alternatively, more sophisticated methods of linkage analysis exploiting the knowledge of a genetic map can be used (Lander & Botstein,1986). Lander and Botstein show that, with a 22cM map (markers separated by intervals of 20%), interval mapping and simultaneous analysis of several markers can reduce the number of families needed to 1/4 of the number required using simple single marker analysis. They also show that the

detection of heterogeneity itself is far more efficient using a map.

In summary, studies of inherited disease have driven the rapid increase in RFLP markers in regions of interest in the genome. As a result, there has been an increase in density of markers all over the genome to such a point that it is now possible to construct genetic maps of chromosomes of reasonably high resolution. These maps will be valuable for the isolation of loci involved in more complex traits and diseases.

Once a cloned genetic marker that displays close linkage to a disease locus has been identified, it may serve as the starting point for the cloning of the gene. This requires an assessment of the physical distance between the two loci in order then to decide on the feasibility of 'walking' (or 'jumping') this distance. The often-quoted approximate ratio of 1cM to 10^6bp is derived from an estimate of 33 Morgans for the sex-averaged genetic size of the genome and an estimate of $3x10^9$bp for the total DNA content (Renwick,1971; Britten & Davidson,1971). The figure of 33M is obtained, not by linkage analysis, but by chiasma counts assuming these to be the cytological counterparts of cross-over events. As well as total counts, chiasma frequency along the lengths of chromosomes can be plotted as has been done for chromosome 1 (Hulten,1974; Hulten et al.1982). Such chiasma maps have been used as a means of relating genetic and

physical distance. This relationship varies along the length of the chromosome. The frequency of chiasmata is smallest at the centromere, increasing distally with a slight decrease in the middle of the arms (Hulten et al.1982).

By comparing the meiotic bivalents with the mitotic physical map, chiasmata can be assigned to chromosomal bands and thereby, genetic distances between bands are estimated. It should be noted, however, that this assumes proportionally equal contraction in meiotic and mitotic chromosomes. Furthermore, there is continuing debate as to whether the positions of chiasmata at diakinesis or metaphase accurately reflect the sites of their formation at diplotene. There is some evidence to support the theory of terminalisation which postulates that chiasmata move distally after diplotene because of the contraction of the chromosomes at diakinesis and their polar reorientation at metaphase. Even neglecting its underlying assumptions, the technique is of limited use; it produces rather coarse results because the physical assignment is limited by the cytological resolution. In the distal regions of chromosomes especially this resolution is far poorer than that achieved genetically.

Thus, until recently the integration of the physical and genetic maps has been limited to the assignment of markers to mitotic bands either by

somatic cell hybrid analysis (Weiss & Green,1967) or by *in situ* hybridisation to metaphase chromosomes (Evans et al.1974). these techniques have allowed the localisation and ordering of markers at the resolution level of chromosomal bands (about 10^7bp).

Pulse-field gel electrophoresis (PFGE- Schwartz et al.1983) allows the separation of DNA fragments approaching 10Mbp. By sequential hybridisation of two closely-linked probes to a single PFGE blot, it is possible to establish a maximum physical distance between the probes if they hybridize to the same fragment. In most regions of the genome (particularly in the distal regions of chromosomes) it is possible to estimate by PFGE the physical distance between markers separated by measurable genetic distances. This technique is obviously limited to cloned markers.

Whilst PFGE has increased the size of DNA that can be resolved by physical techniques, the application of the polymerase chain reaction (PCR) to single cells may allow the measurement of much smaller genetic distances (Boehnke et al.1989). Single sperm cells represent individual meiotic events which, unlike offspring from informative parents, are essentially unlimited. thus, if sperm from an individual doubly heterozygous for two closely-linked markers is obtained and the haplotype of a large number of sperm determined, then the frequency of rare recombination events (i.e. short genetic distances) may be measured. PCR-sperm analysis also

permits the investigation of environmental factors on recombination (e.g. age) since recombination fractions can be obtained from a single individual. The obvious limitation of this technique is that only male recombination can be studied.

The investigation of the relationship between physical and genetic distance can now be carried out at the molecular level. Indeed, the validity of the concept of genetic distance for very short physical distances in the context of mapping may be questioned. Genetic distance is a useful measurement for, for instance, the estimation of the distance needed to be covered between a disease locus and an RFLP marker. For very short physical distances, however, the assumption that recombination is random may not be valid. The presence or absence of recombinational hot-spots in a particular segment of DNA would obviously have a large bearing on genetic distance.

Obvious regions to study these questions are the distal regions of chromosomes where the frequency of recombination and thus the chance of physically linking genetically-separated probes is high. Physical maps of the pseudoautosomal region have already been constructed (Brown,1988; Petit et al.1988). These have shown that this region which by definition covers 50cM is about 2500kb long suggesting, as expected, that this region is recombinationally very active. Furthermore, both papers reported that recombination sites were not

markedly clustered. The distal region of the short arm of chromosome 1 is another region which displays high recombination frequencies. 1p36-1pter, which includes the PGD locus (Carritt et al.1982), represents about 20% of the physical length of 1p (Paris Conference,1971). However, chiasmata counts of meiotic chromosomes suggest that 1p36 accounts for about 50% of the genetic length of the male meiotic chromosome 1. This technique cannot provide any estimate of female genetic distances. Several RFLP maps of this region have recently been published (Dracopoli et al.1988; O'Connell et al.1989; Donis-Keller et al.1989). Genetic mapping of other chromosomes has indicated that although female rates of recombination are generally higher than those in males, this difference is diminished and even reversed in the distal regions of chromosomes (White et al.1985; Lathrop et al.1988). However, neither O'Connell et al. (1987) or Dracopoli et al. (1988) have found any evidence for this on chromosome 1p.

Interest in this region of the genome has been stimulated by its involvement in a number of tumours. Loss of heterozygosity at loci in the distal region of 1p has been observed in neuroectodermal tumours including MEN2 (Mathew et al.1987), neuroblastoma (Fong et al.1989) and malignant melanoma (Dracopoli et al.1989) as well as in testicular teratomas (Parrington et al.1987).

The frequent but inconsistent involvement of 1p in different tumours suggests genes in this region may play a role in tumour development rather than initiation.

PGD has not been implicated in any of these tumours. Interest in PGD stems from its location in the distal region of chromosome 1p. The polymorphism exhibited by the PGD protein was used extensively for the initial genetic mapping of chromosome 1. However, the informativeness of this polymorphism is low relative to the majority of RFLPs. It was one of the principle aims of this project to increase the genetic information detectable at this locus by cloning the gene and identifying a useful RFLP.

The development of pulsed-field electrophoresis has now made it possible to estimate the physical separation between genetically distinguishable loci The availability of a DNA probe for the PGD locus would therefore allow a more precise investigation of the relationship between genetic and physical distances in this region of chromosome 1p.

Finally, the availability of genomic clones of the PGD gene (particularly the 5' upstream sequences) for structural and sequence analysis would certainly help elucidate the molecular mechanisms underlying the dietary and hormonal influences on PGD expression. Sequence comparisons may also shed light on the evolutionary relationships between the enzymes of the pentose phosphate pathway.

CHAPTER 2 METHODS AND MATERIALS

2.1 DNA Isolation

2.1.1 Large-Scale Isolation of Plasmid DNA.

5ml of L-broth containing 50µg/ml of ampicillin was seeded with a single colony scraped from an agar plate with a sterile loop and incubated in a 100ml flask for 5-16 hours at 37°C with vigorous shaking (200rpm) to ensure adequate aeration of the culture to encourage aerobic growth. 250µl of this culture was in turn used to seed 250ml of L-broth in a 1l flask (again containing 50µg/ml of the appropriate antibiotic solution) and this was grown overnight under the same conditions.

The following day the cells were harvested by centrifugation at 6000rpm for 10min at 4°C (in an MSE Europa24M centrifuge). The supernatant was poured off and then the remaining broth was removed by aspiration. The procedure of Birnboim and Doly (1979) was followed for the subsequent isolation of the plasmid DNA from this cell pellet with minor modifications. The cell pellet was resuspended in 10ml of 'lysozyme buffer' (without lysozyme- 25mM Tris.Cl,pH8.0, 50mM glucose, 10mM EDTA). The cells were then lysed by adding 20ml of alkaline SDS (0.2N NaOH/1% SDS). Genomic DNA was removed by precipitation with 15ml of 'acid potassium

acetate' (3M KOAc/11.5% glacial acetic acid) followed by centrifugation at 10,000rpm for 15min at 4°C. The plasmid DNA was precipitated from the supernatant by mixing the latter with 0.6 volumes of propan-2-ol. This was left at room temperature for 15min to minimize the precipitation of RNA. The precipitate was collected by centrifugation at 10,000rpm for 10min. The precipitate was washed with 70% ethanol before drying and dissolving in 11.0ml of TE for centrifugation through a caesium chloride-ethidium bromide gradient.

When the pellet was fully dissolved 0.44ml of 0.2M K_2PO_4 and 11.44g of CsCl were added and when these were dissolved, 1.14ml of ethidium bromide (10mg/ml) was added. This solution (1.55g CsCl/ml) was transferred to a Sorvall sealable centrifuge tube. The tube was filled with paraffin oil, sealed and spun at 45,000 rpm at 20°C for 16hrs in a Sorvall OTB75B ultracentrifuge.

The plasmid band was removed through a needle by syringe, extracted with CsCl-saturated isoamyl alcohol and dialyzed against 4-8 litres of TE. Finally, the DNA was precipitated in 70% ethanol first in 0.4M NaCl and then in 3.75M NH_4OAc. Plasmid DNA was stored in TE at -20°C.

2.1.2. Small-scale Isolation of Plasmid DNA

The initial steps of this method are essentially the same as those for the large-scale isolation

procedure. Cells are collected from 1.5ml of an overnight culture, suspended in 100µl of lysozyme buffer, lysed with 200µl of alkaline SDS and the genomic DNA precipitated with 150µl of acid potassium acetate. The supernatant containing the plasmid DNA was extracted once with TE-saturated phenol. The aqueous layer was than precipitated at room temperature (for 2-5 minutes) with two volumes of ethanol. The DNA precipitate was recovered by centrifugation in a minifuge (13,000 rpm) for 10 minutes. The DNA was redissolved in 20µl of deionised water. Subsequent retriction digests of DNA prepared in this way required spermidine (5mM) and RNaseA (1-5µg/reaction) because of the presence of contaminants including RNA.

2.1.3. Small-scale Isolation of Single-stranded Bacteriophage M13 DNA for Sequencing

1.5ml of 2xTY medium with 15µl of an overnight culture of JM101 (Messing et al.1981) was innoculated with a single M13 plaque (colourless). The culture was incubated at 37°C for 4-5 hours with vigorous shaking. All subsequent steps were performed at room temperature and centrifugation was carried out at 13000rpm in a minifuge. The cells were pelleted by spinning for 5 minutes. The pellet (containing the replicative double-stranded form of the phage) was discarded and the supernatant was re-spun for 5 minutes and any remaining cells again removed. The phage particles were precipi-

tated by adding 200µl of 20% PEG/2.5M NaCl, vortexing and allowing to stand at room temperature 15-60 minutes. The phage were spun down for 5 minutes and the supernatant carefully aspirated off (contaminating PEG interferes with subsequent sequencing of the DNA). 100µl of TE and 50µl of TE-saturated phenol was added and the mixture vortexed thoroughly. After separation of the phases by centrifugation for 3 minutes, the aqueous layer was precipitated at -20°C with 1/10 volume 3M sodium acetate,pH5.5 and 2 volumes ethanol. The DNA was collected by centrifugation for 10 minutes and dissolved in 30µl TE. The yield was usually 5-10µg. 1-2µg was used for each sequencing reaction.

2.1.4. Large-scale Isolation of Phage DNA

A single colony of the *E.coli* host was grown overnight in the appropriate medium. 1ml of this culture was used to seed 100 ml of medium with 10mM $MgSO_4$ in a 1l flask. The culture was incubated at 37°C (with vigorous shaking) until a cell density of about $1x10^8$/ml ($O.D._{600}$ ~ 0.2) was reached. The culture was then infected with $1-5x10^5$ pfu of a phage lysate and allowed to incubate at 37°C without shaking for 15min. The culture was then grown a further 4-6 hours until lysis was observed. Lysis was completed by mixing with 1ml of chloroform. The cell debris was removed by centrifugation at 7500rpm at 4°C for 30min. Phage particles were precipitated by dissolving 7.5g of poly-

ethylene glycol (PEG) 6000 and 6g of NaCl in the supernatant and allowing it to stand for at least 1 hour at 4°C. The precipitate was collected by centrifugation at 7500rpm at 4°C for 15min. The pellet was suspended in 5ml phage buffer prior to purification on a step caesium chloride gradient as described by Maniatis et al.(1982).

2.1.5. Isolation of Genomic DNA for Southern Analysis

From blood: 20ml of whole blood was spun at 2500rpm for 10min at 4°C (in a Heraeus swing rotor centrifuge). The white cell layer was removed with a pasteur pipette. If the DNA was not immediately isolated, the cells were stored at -70°C. Frozen samples were thawed and suspended in 4ml of autoclaved deionised water in order to lyse the contaminating red cells. Intact white blood cells spun down at 2500rpm at 4°C for 15min and gently resuspended in 4ml STE (see Materials) with 2mg of proteinase K and 0.2ml of 10% SDS. Lysis was allowed to proceed overnight at 37°C. The digest was then extracted twice with phenol saturated with TE by gentle mixing to minimize DNA shearing. Contaminating phenol was removed by extraction with 4ml of chloroform/isoamyl alcohol (24:1). Each time the aqueous layer was removed with a wide-bore pipette. This was then precipitated with 2 volumes of ethanol. The visible precipitate was 'hooked out' with a sealed, bent pasteur pipette. 1ml of TE was added and

the DNA allowed to redissolve at 4°C (for 16-48 hours). When dissolved the DNA was reprecipitated in 70% ethanol/3.75M ammonium acetate. This final precipitate was redissolved in 0.2-0.5ml TE and stored at -70°C.

From tissue culture cells: About 10^7 confluent cells in a large flask ($150cm^2$) were washed twice in 0.9% saline. The cells were then overlayed with 10ml STE with 200µg/ml proteinase K and 0.5% SDS and incubated overnight at 37°C. The lysed cells were then extracted with phenol and the DNA precipitated as described above.

2.1.6. Isolation of High Molecular Weight Genomic DNA for Pulsed-Field Gel Electrophoresis

This method is based on that of Schwartz and Cantor (1984). Confluent tissue culture cells were collected by trypsinisation (see Tissue culture techniques, section 2.10) and washed twice in 0.9% saline solution. The cells were counted in a haemocytometer and the density abjusted to $1x10^7$ cells/ml. This suspension was mixed with an equal volume of 1% low melting point agarose (InCert from FMC Bioproducts), aliquotted into 100µl moulds and allowed to set. 100 blocks were then placed in 50ml proteinase buffer (1% sodium lauroyl sarcosinate (Sigma)/0.5M EDTA/2mgml^{-1} proteinase K) for 48 hours at 50°C. The buffer was renewed after 24 hours. The blocks were then washed

extensively in sterile TE at 50°C (four times). The proteinase K was then inactivated by washing the blocks in 50ml TE with 0.04mg/ml phenylmethylsulfonylfluoride (PMSF, prepared as a 40mg/ml solution in propan-2-ol). The blocks were then washed for a final time in 2x50ml of TE and then stored in 0.5M EDTA,pH8 at 4°C. The condition of the DNA was assessed by running it undigested on a pulsed-field gel.

2.2 Isolation of RNA

All materials and solutions used in the isolation and analysis of RNA which did not contain Tris were treated with Diethylpyrocarbonate (DEP) in order to minimize degradation of RNA by RNAse A activity. DEP was added to solutions to a final concentration of 0.1% and the solutions were incubated at 37°C overnight. Solutions were then autoclaved to inactivate the DEP. Plastics were similarly treated with 0.1% DEP solution after silicionisation with dimethyldichlorosilane solution.

2.2.1 Isolation of Total RNA

The method used was modified from Chirgwin et al.(1979). Tissue culture cells were collected by centrifugation (after trypsinisation and scraping in the case of EJ cells) and washed twice in 0.9% sterile saline solution. The cells were counted on a haemocyto-

meter and 2-5x10^7 cells lysed by vortexing in 10ml of guanidium chloride solution (6M GuCl/0.2M NaOAc,pH5.6). Whilst vortexing, 5ml of ethanol (BDH Aristar) was added and the RNA allowed to precipitate at -20°C overnight. This was collected by centrifugation at 10,000rpm for 15 minutes at 4°C. The pellet was dissolved in 2ml urea solution (7M urea/0.35M NaCl/1mM EDTA/0.2% SDS/50mM Tris.Cl,pH7.5). When the pellet was fully dissolved, the solution was extracted with an equal volume of 1:1 phenol/chloroform. The aqueous layer was precipitated with 2 volumes of ethanol, the pellet washed with 70% ethanol and finally redissolved in 100-500µl DEP-treated deionised water.

2.2.2 Isolation of Poly(A)$^+$ RNA from Total RNA

Poly(A)$^+$ RNA was recovered by oligodT-chromatography of total RNA based on the protocol of Aviv and Leder (1972). 600mg of oligodT-cellulose was equilibrated with 30-40ml of HS solution (0.5M NaCl/0.1% SDS/10mM Tris.Cl,pH7.4) in a Whatman chromatography column kept at 40°C by means of a water-jacket. The column was operated with gravity feed at a flow rate of 1-2ml/min.

10-20mg of total RNA was dissolved in 5ml of HS solution and run through the column. The eluate was recycled three times and finally washed through with 30ml of HS or until the eluate $O.D._{600}$ was less than 0.05. The poly(A)$^+$ RNA was eluted in 15ml of LS

solution (0.1% SDS/10mM Tris.Cl,pH7.4). The eluate was collected until the $O.D._{600}$ was less than 0.05. This eluate was adjusted to 0.5M NaCl with 5M NaCL solution and the entire process was repeated. The $O.D._{600}$ of the final LS eluate was determined before the RNA was precipitated with 1/10 volume sodium acetate (2M,pH5.5) and two volumes of ethanol.

The poly(A)$^+$ was stored as an ethanol precipitate until required when an appropriate amount was spun down and redissolved in deionised water.

2.3. Gel Electrophoresis and Blotting

2.3.1 DNA Agarose Gel Electrophoresis

Analytic cloned DNA digests (with 0.5-1.0µg) were performed in 20µl of the recommended 1xBRL restriction enzyme buffer (with added bovine serum albumin-50µg/ml and spermidine-5mM) for 1-3 hours at the appropriate temperature. Preparative DNA digests were performed in 200µl reaction volumes (with 10-30µg of DNA). After digestion the DNA was precipitated in 0.4M NaCl/70% ethanol and redissolved in 25µl of deionised water before loading. Genomic DNA digests were performed in 30ul of the appropriate 1xBRL buffer (with added BSA and spermidine) for 3-6 hours.

Routine analytic and preparative gels were 0.7-1.4% agarose dissolved in 100ml 1xTAE and run in 1xTAE buffer (with 0.5µg/ml ethidium bromide in both the gel

and running buffer). Gels run at more than 2V/cm were made and run in 1xTBE. DNA fragments prepared for radiolabelling were run in 0.7-1.0% low melting point agarose.

2.3.2 Southern Transfer

These were performed essentially as described by Southern (1975) except that the DNA was depurinated prior to denaturation by placing the gel in 0.25N HCl for 2x5min. This improves the efficiency of the subsequent transfer (Wahl et al.1980). Genomic gels were blotted for three hours. Gels containing cloned DNA were routinely blotted twice (first for 20-30min, second for 40-90min). Transfer was carried out in 10xSSC onto Genescreen Plus (NEN Research Products) membranes. Membranes which were to be probed more than once were cross-hybridized by exposure to 260nm U.V.light for 2min.

2.3.3 Pulse-Field Gel Electrophoresis

Generally, half blocks ($5x10^5$ cells, about 2-3µg of DNA) were digested and loaded into each well. These blocks were removed from the 0.5M EDTA,pH8 storage solution and rinsed twice in sterile TE. The blocks were then placed in the appropriate 1x BRL restriction buffer with 5mM spermidine and 50µg/ml BSA (total volume including block=200µl). 10-20 units of

restriction enzyme was added and the reaction incubated on ice for 30 minutes to allow the enzyme to diffuse into the block before digestion at 37°C for 4-16 hours.

In the case of double digests with NarI and another enzyme, the NarI digestion was performed first in BRL buffer #1 (no NaCl), then the NaCl was adjusted and the second enzyme was added. Mock digests without any restriction enzyme were always performed in parallel in order to assess the level of non-specific DNA degradation during digestion. After digestion blocks were loaded into the gel and sealed in with molten 0.8% low melting point agarose. Yeast chromosomes and lambda DNA oligomers in blocks (from FMC BioProducts) were used as size markers.

Pulse-field gel electrophoresis (PFGE) was carried out in LKB-Pharmacia Pulsafor apparatus using a hexagonal array electrode. 150ml gels (0.8-1.2%) were made in 0.5xTBE. The running buffer was also 0.5xTBE. Electrophoresis was performed in the absence of ethidium bromide. Gels were run at 170V (field strength= 6V/cm) at 4°C for 24-40 hours. Pulse times used were between 20 seconds (resolution range: 40-300kb) and 125 seconds (resolution range: 200-1600kb).

2.3.4 Transfer of PFGE DNA to Genescreen Plus

PFE gels were blottedin the same way as standard DNA gels except that thay were not depurinated with

acid prior to denaturation. Instead gels were exposed to 260nm U.V. light for 8 minutes in order to fragment the DNA. Gels were blotted overnight and filters were cross-linked prior to hybridisation. Filters were hybridized in identical fashion to Southerns (see section 2.5.1).

2.3.5 RNA Agarose Gel Electrophoresis

The RNA electrophoresis protocol is modified from Thomas (1980). 10-20µg of total RNA or 3-5µg of poly(A)$^+$ RNA was incubated in 1M glyoxal/50% DMSO/0.01M phosphate buffer, pH7.0 at 50°C for 1 hour prior to loading. The RNA was run at 100V for 3 hours on a 1.4%/0.01M phosphate buffer,pH7.0 agarose gel. The running buffer (also 0.01M phosphate,pH7.0) was recirculated during the run to ensure even temperature distribution. Gels were not routinely stained but if so they were immersed in 0.5µg/ml ethidium bromide for 30-60min.

2.3.6 Northern Transfer

Transfer was performed as described by Thomas (1980). The RNA was transferred overnight to Genescreen Plus membranes in 10xSSC.

2.4 Library Screening

The genomic and cDNA libraries were screened by the method of Benton and Davis (1977) with minor modifications. In the case of λgt10 cDNA libraries, the host cells used were *E.coli* NM514 and the Charon4A genomic library host was *E.coli* LE392. In both cases, the libaries were plated out (as described in section 2.9) onto 20cmx20cm sterile plastic culture plates (Nunc) and incubated overnight. About 10^6 plaques were plated out on two such plates. The plates were kept at 4° C prior to lifting.

Nitrocellulose filters (Schleicher & Schuell) were soaked in sterile 3xSSC and then placed on the phage plate for 90 seconds during which the position of the filter was marked with a needle. The filter was then placed phage side up on 3MM filter paper soaked in denaturing solution (0.5N NaOH/1.5M NaCl) for 3 minutes. The filter was then transferred successively to two baths of neutralizing solution 0.5M Tris.Cl,pH7.4/3M NaCl (for 1 and 3 minutes). The filter was air-dried before baking at 80° C for 2-3 hours. All lifts were carried out in duplicate. The second filter was left in contact with the phage plate for 3 minutes. After hybridisation of the filters (see section 2.5.2) and autoradiography positive plaques were picked (see section 2.9) and taken through two more rounds of plating and screening.

2.5 Radiolabelling of Probe DNA

2.5.1 Random-Primer Labelling

DNA fragments were labelled by primer-directed labelling (Feinberg & Vogelstein,1984).

Probe DNA was separated from vector DNA by restriction enzyme digestion followed by electrophoresis through low melting point agarose. The gel slice containing the DNA was removed and weighed. Three times this mass of water was added and gel slice dissolved at 100°C for 7min. Aliquots containing 30-50ng of denatured DNA were stored at -20°C.

DNA was labelled using the Multi-prime labelling kit (Amersham) as directed by the manufacturer This is based on the method of Feinberg and Vogelstein (1984). Essentially, small random oligonucleotides (6mers) are annealed to the fragment to be labelled and extended by the Klenow fragment of *E.coli* DNA polymerase in the presence of ^{32}P-dCTP. The radioactive products are short copies of the original target fragment. Probe DNA was separated from unincorporated labelled nucleotides by centrifugation through Sephadex G-50.

2.5.2. End-Labelling

The products of the HpaII digestion of P1.4 (section 3.11, fig.18) and the markers were end-labelled by the method of Drouin (1980) with Klenow fragment of *E.coli* DNA polymerase (BCL) and ^{32}P-dCTP.

This will label HpaII fragments since the HpaII produces 5' protruding GC ends. Thus Klenow will add a single ^{32}P-dCTP nucleotide to each fragment end. 2 units of Klenow and 10µCi of the labelled nucleotide were added to the DNA digests (still in their restriction enzyme buffer) and incubated at room temperature for 30 minutes. Unincorporated label was removed as before by spin chromatography through sephadex G-50.

2.6 Filter Hybridisation

2.6.1 Genescreen Plus Filters

Genescreen Plus filters (NEN Research Products) were hybridized in plastic bags as directed by the manufacturer's protocol. Briefly, filters were pre-hybridized in 10ml of 10% dextran sulphate/1M NaCl with 1-2% SDS for at least 15min. Cross-linked filters were pre-hybridized in 2% SDS and untreated filters in 1% SDS. Labelled DNA was denatured with 250µl of sonicated salmon sperm DNA (5mg/ml) at 100°C for 5min before being added to the the pre-hybridisation solution. Southern filters were hybridized at 65°C and Northern filters at 60°C for 16-20 hours.

Genescreen Plus Southern blots (and PFGE blots) were initially washed in 2xSSC at room temperature (2x10min) and then in 2xSSC/1%SDS at 65°C (2x30min). Moderate stringency washes were then completed in

0.1xSSC at room temperature (2x15min). High stringency washes were carried out in 0.1xSSC at 65°C (2x15min). Northern filters were washed in the same fashion except at 60°C rather than at 65°C where appropriate. Filters were wrapped in polythene prior to autoradiography to prevent dehydration.

2.6.2 Stripping Probe from Genescreen Plus

Genescreen plus membranes which had previously been crosslinked the the DNA after Southern transfer were stripped by placing them in 0.1xSSC/1% SDS for 10-20 minutes as recommended by the manufacturer. Stripped filters were autoradiographed for at least 16 hours prior to rehybridisation to ensure that no signal remained on the filter. This is of crucial importance with PFGE blots if physical linkage between the two probes is being sought.

2.6.3 Nitrocellulose Filters

Nitrocellulose filters were pre-hybridized in 5x Denhardt's/5xSSC for at least 5 hours. Labelled probe was added and hybridisation allowed to proceed for 16-20 hours.

Filters were initially washed in 2xSSC at room temperature (2x15min) and then at 65°C in 2xSSC/0.1%SDS (2x30min). Final washes at moderate stringency were performed in 0.1xSSC (2x15min) at room temperature or

at high stringency in 0.1xSSC at 65°C (2x15min). Filters were allowed to air-dry before autoradiography.

2.7 In Situ Hybridisation

2.7.1 Fluorescent *In Situ* Hybridisation

Fluorescent *in situ* hybridisation of biotinylated pPGDH4 was carried out by D.Griffin, Galton Laboratory.

Biotinylated probe was prepared by labelling whole plasmid by nick translation using a BRL kit. Incorporation of label was assessed by using the BRL streptavidin-alkaline phosphatase kit.

The fluorescent in situ hybridisation protocol used was adapted from Pinkel et al.(1986) as described in Griffin et al.(submitted).

Essentially, metaphase chromosomes were G-banded by treatment with 2xSSC and trypsin. Slides were destained and dehydrated in an alcohol series and dried. They were pre-treated with RNase (100µg/ml in 2xSSC/37°C/1 hour) in a moist chamber. After proteinase K treatment (50µg/ml/37°C/7 minutes), the slides were fixed with 4% paraformaldehyde, washed and dehydrated. The hybridisation buffer was 50% formamide, 20% dextran sulphate, 50mg/ml denatured salmon sperm DNA, 1mM EDTA in 2xSSC. 20ng of probe was used per slide.

After simultaneously denaturing the probe with the target DNA at 80°C/4 minutes, hybridisation was allowed

to proceed overnight. The slides were washed in 50% formamide/2xSSC (3x5min.) and then in 2xSSC (5x2min.). Biotin detection was facilitated by 2 sequential layers of Avidin-FITC (fluorescein isothiocynate) conjugate and biotin-anti avidin D conjugate (Vector Laboratories,USA). Final washes (5minutes each) were in 0.05% Tween 20/4xSSC and twice in 0.9% phosphate-buffered saline. Slides were observed after treatment with 'anti-fade' medium (DABCO) under a Reichert polyvar microscope fitted with ultraviolet fluorescence and filters for DAPI and FITC.

2.7.2 Tritium *In Situ* Hybridisation

In Situ hybridisation with tritiated pPGDH1 was carried out by K.Gulati, Galton laboratory.

The whole plasmid was nick translated using a BRL kit. Incorporation was assessed using a scintillation counter. Metaphase chromosomes were lipsol-banded. *In situ* hybridisation was performed as described in Harper & Saunders (1981).

In essence, the slides are RNase-treated, denatured (in this case the probe is denatured separately from the target DNA), and hybridized in 50% formamide/5x Denhardts/5xSSPE with 10ng probe/slide at 42° C overnight. The slides are then washed down to 0.1xSSC, dehydrated and air-dried. They are then dipped in emulsion (Ilford K2) and stored for 2-4 weeks in a

lightproof box before developing , giemsa staining and viewing under oil immersion.

2.8 DNA Cloning

2.8.1 DNA Fragment Preparation

Restriction digests were run in agarose gels and the DNA fragments to be cloned were run onto DEAE (45µm) membranes (Schleicher & Schuell). DNA was eluted by immersion in 500µl 2M NaCl at 70°C for 15-30 minutes. This solution was then dialysed against 2-4l of TE (in a Sartorius dialysis apparatus) before precipitation in 70% ethanol/0.4M NaCl. DNA concentration was assessed by visualisation in an ethidium bromide agarose gel with standard concentration DNA markers.

Vectors used were the general purpose vector, pAT153 (Twigg & Sherratt,1980), a derivative of pBR322 and MI3mp18 (Yanish-Perron et al.1985) for single-stranded sequencing. These were digested with the appropriate cloning site enzyme(s). The digest was then extracted once with TE-saturated phenol and once with chloroform. If the two cloning sites were identical, these were phosphatased with calf intestinal phosphatase using the protocol of Maniatis et al.(1982). After inactivation at 65°C, the digest was again extracted once with phenol and once with chloroform.

2.8.2. DNA Ligation

Ligations were performed in 25-50µl 1xBRL ligation buffer with 5 units of BRL T4 DNA ligase. Ligations were performed at 15°C for 3-16 hours. Molar ratios of vector DNA to insert DNA were between 1:1 and 5:1 with the total mass of DNA between 50-200ng. Control ligations with no insert DNA were always carried out in parallel.

2.8.3. Transformation of Competent *E.Coli* Cells.

The cells used were *E.coli* 831 as pAT153 hosts and JM101 (Messing et al.1981) as M13mp18 hosts. 831 are a stably-recA$^-$ strain (C.Gilliam, personal communication) that were used as hosts for all plasmids containing human PGD cDNA and genomic fragment plasmids. Both these cells were made competent and transformed using a simple calcium chloride method essentially as described by Mandel and Higa (1970). Briefly, an overnight culture of cells was used to seed a 50-100ml culture which was grown to an $O.D._{600}$ of 0.3-0.5 (831) or $O.D._{600}$ of 0.3-0.4 (JM101). The cells were recovered by centrifugation and resuspended in half the original volume of transformation buffer (10mM Tris.Cl,pH8/50mM $CaCl_2$) and incubated on ice for 15-30min. The cells were again recovered and resuspended

this time in 1/15 volume transformation buffer. They were either stored for 24 hours at 4°C before transformation or else glycerol was added to a final concentration of 15% and they were stored at -70°C in 200µl aliquots for later use. Cells stored in this fashion retained an acceptable level of competency for 3-6 months. Competency was assessed by transformation with a known amount of pAT153 (10-30ng). Transformation efficiency varied between 10^5-10^6 colonies per µg of plasmid.

Competent cells were incubated with 20-50ng of ligated DNA on ice for 30 minutes. Cells were then heat-shocked at 42°C for 2 minutes, cooled on ice and finally incubated at 37°C in 2ml L broth for 60 minutes to allow expression of antibiotic resistance. The cells were plated on a Millipore triton-free nitrocellulose filter on L agar containing the appropriate antibiotic.

All the human cDNAs (pPGDH1-4) and genomic fragments (pPGDE4 and E10) were subcloned into the EcoRI site of pAT153 and so the antibiotic used to select transformed cells was ampicillin (50µg/ml final concentration).

2.9. DNA Sequencing

2.9.1. Sequencing Reactions

DNA fragments to be sequenced were cloned into the EcoRI site of M13mp18 as described in section 2.6. The

ssDNA form of the selected clones was isolated as described in section 2.1.3. Since the frequency of false positives using the Xgal/IPTG selection system for recombinants was often very high, batches of ssDNAs were T-tracked (only the sequence of T residues is determined) prior to determination of the complete sequence in order to identify true recombinants and complementary pairs of clones with inserts in opposing orientations. DNA sequencing was carried out using the Sequenase 2.0 kit (USB) and the manufacturers instructions were followed exactly. This is a dideoxy nucleotide chain termination method based on that of Sanger et al.(1977). In essence, ssDNA samples mixed with primer complementary to a region upstream of the polylinker site in M13. After brief heating to disrupt secondary structures, the two strands are allowed to anneal by slow cooling. The primer is then extended in the presence of dGTP, dCTP, dTTP and ^{35}S-labelled dATP by a modified T7 DNA polymerase (Sequenase 2.0). This reaction is aliquotted into 4 tubes each containing the four deoxy nucleotides and one of the four dideoxy nucleotides at a tenth of the concentration of the deoxy nucleotides. The dideoxy nucleotides cause chain termination when they are incorporated. The reaction was completely stopped by the addition of 4µl of 95% formamide/20mM to 6µl reaction mixtures. T-tracking was performed in identical fashion to sequencing except that initial amounts of primer and DNA were reduced to

25% and all of the extension reaction was transferred to a single tube with the termination mix containing dTTP. The DNA was denatured briefly before loading onto a sequencing polyacrylamide gel.

2.9.2. Polyacrylamide Gel Electrophoresis of Sequencing Products

A Bio-Rad 21x50 Sequi-Gen apparatus was used for electrophoresis.

Gels were prepared and poured as described in the manufacturer's protocol. Pre-mixed 19:1 acrylamide:bis-acrylamide (Bio-Rad) was used in order to minimize exposure to unpolymerized acrylamide powder. Simple two-step gradient denaturing gels were poured by preparing two gel mixes: The 'heavy' mix was 2.5xTBE/6% acrylamide/7.67M urea/5% sucrose and the 'light' mix was 0.5xTBE/6% acrylamide/7.67M urea. The bottom 13.5ml of the gel was a 1:1.5 light to heavy mixture whilst the remaining 28.5ml of the gel was light. Polymerisation of the gel with ammonium persulphate and TEMED was adjusted such that the gel was set in 15-30 minutes.

At least two hours after pouring, the gel was pre-run at 1800-2200V in 1xTBE until at 40-50°C before loading the samples with a shark-tooth comb. Samples were run 1800-2200V for 3-6 hours or at 1000V for 16-20 hours.

After dismantling the apparatus the gel was removed from the glass plate onto 3MM filter paper (Whatman) and then dried down on a BioRad 1125B vacuum gel drier at 80°C. After being allowed to air-dry completely at room temperature, the gel was exposed to autoradigraphic film for 24-72 hours. Sequence was read by eye and analyzed using the Seq programs from Intelligenetics.

2.10 Starch Gel Electrophoresis

PGD enzyme electrophoretic phenotypes were determined by starch gel electrophoresis as described by Harris and Hopkinson (1976). 15cmx22cm starch gels were 11% in 0.01M phosphate buffer,pH7.0. The running buffer was 0.1M phosphate buffer,pH7.0. The inclusion of NADP (20µg/ml) in both the gel and running buffer improved results. The gels were run at 4°C for 24 hours at 50V (the PGD enzyme runs anodally).

The gels were stained (after slicing) as described by Harris and Hopkinson (1976) using the PMS/MTT detection system. The following mixture was placed on top of the interior face of one half of the gel: 10ml 0.5M Tris.Cl,pH8, 10mg 6-phosphogluconate, 5ml og 0.2M $MgCl_2$, 1ml of 5mg/ml NADP, 1ml of 5mg/ml MTT, 1ml of 5mg/ml PMS and 20ml of 2% agar. The staining was allowed to occur for 10-40 minutes at 4°C. The dehydrogenation of 6-PG by PGD reduces NADP to NADPH which in

turn reduces PMS which finally causes the reduction of MTT (yellow) to formazan (blue). All gels were run with AA and AB controls.

2.11 Bacterial Cell Culture Techniques

2.11.1. Bacterial Cell Growth

E.coli 831 cells were initially streaked on L-agar and then grown in L-broth (see Materials). All human DNA was cloned into the EcoRI site of pAT153 and thus *E.coli* carrying these plasmids were grown on agar or in L-broth containing 50µg/ml ampicillin (Sigma). JM101 were initially streaked on minimal plates (M9) and then grown in 2xTY.

2.11.2. Preparation of Bacteriophage Host Cells

E.coli LE392 (Borck et al.1976) and NM514 (Frischauf et al.1983) cells used for plating out Charon4A and λgt10 libraries respectively were both grown in L-broth without glucose but with 0.2% maltose. This induces the maltose operon which includes *lamB*, the phage attachment site. Overnight cultures of these cells were spun down and resuspended in the original culture volume of phage buffer (10mM $MgCl_2$/10mM Tris.Cl,pH7.6) and stored at 4°C. They were kept for up to 2 weeks.

2.11.3. Bacteriophage Plating

L-agar plates were poured (on sterile Nunc plates) and allowed to dry in a hood. For 20cmx20cm plates (200ml agar), 200µl of host plating cells (prepared as described above) and 200µl of sterile 10mM $MgCl_2$/10mM $CaCl_2$ were mixed with an appropriate amount of phage stock. The mixture was incubated for 15 minutes at 37°C before adding 40ml of slightly cooled (~50°C) molten top agarose (see Materials). This was quickly poured over the agar plate. The top agarose was allowed to set (5 min) and then the plates were incubated overnight at 37°C. For subsequent rounds of library screening phage were plated out on 9cm plates (20ml agar and 3ml top agarose) in an identical way except that half the amounts of plating cells (100µl) and Mg^{2+}-Ca^{2+} (100µl) was used. Plating of M13 bacteriophage was carried out in exactly the same way except that H-agar and H-top agar were used instead. Also, 40µl of IPTG (100mmol) and 40µl of Xgal (2%) were included in the H-top agar. This allows the selection of recombinant clones which are clear (rather than blue) because of the disruption of the lac operon in M13mp18.

2.11.4. Phage Stocks

Phage stocks were prepared from confluent 9cm plates prepared as above. The phage were eluted into 5ml phage buffer (10mM $MgCl_2$/1omM Tris.Cl,pH7.6) pipetted

onto the plate by gentle rocking for at least 2 hours. The stocks were kept at 4°C with a drop of chloroform to prevent bacterial cell growth.

2.12. Mammalian Cell Culture

Cell lines were grown in the following media:

MGL8B2:	RPMI 1640 + 10% foetal calf serum (FCS)
EJ:	Dulbecco's modified MEM
a23/Faza:	MEM
DIS2.6:	Eagle's Medium: MEM + 10% FCS + non-essential amino acids
All other hybrids:	F12 + HAT

2.13 Somatic Cell Hybrid References

DIS	Carritt and Povey, 1979.
Taxi	Carritt et al.1982.
HCH	Carritt et al.1982.
C4 & B4	Carritt (1980).
HORL	Van Heyningen et al.1975.
CTP	Jones et al.1976.
FHA7	Kielty et al.1982.

2.14 Materials

All chemicals were obtained from BDH Chemicals Ltd., Sigma Chemical Co. or Pharmacia Ltd.

Cell culture reagents were purchased from Gibco(UK) Ltd., Flow Laboratories and Difco Laboratories (USA).

Restriction enzymes were obtained from Bethesda Research Laboratories (BRL), New England BioLabs and

P&S Biochemicals.

2.14.1 Stock solutions

1. SSC solutions: 1XSSC is 0.15M sodium chloride
   0.015M sodium citrate

   All SSC solutions were pH 7.0-7.5.

If not, it was adjusted with citric acid.

2. TE: 1mM Tris.Cl, 0.1mM EDTA,pH8.0

3. STE: 150mM NaCl, 10mM Tris.Cl,pH8.0, 10mM EDTA

4. TAE electrophoresis buffer.

   Made up as 5xTAE: 0.2M Tris.Cl,pH8.0
   0.1M sodium acetate
   5mM EDTA

5. TBE electrophoresis buffer.

   Made up as 5xTBE: 0.45M Tris,pH8.2
   0.45M boric acid
   10mM EDTA

6. Phosphate buffer.

Adjusted one with the other to pH7.0. RNA gel electrophoresis buffer was a 1:10 dilution treated with DEP (0.1%) and autoclaved. Starch gel electrophoresis running buffer was a 1:100 dilution and the gel buffer was a 1:10 dilution.

7. Salmon sperm DNA

Salmon sperm DNA was dissolved at 5-10mg/ml in deionised water at 65°C. It was then boiled for 5 minutes, quenched on ice and sonicated for 10 minutes

(MSE Soniprep 150). Finally the concentration was adjusted to 5mg/ml and it was stored at -20°C.

8. Denhardt's Solution

   Made up as 20x: 4g/l BSA
   4g/l PVP360
   4g/l Ficoll
   1.25g/l salmon sperm DNA
   (prepared as above)
   0.4M phosphate buffer

9. Phage buffer

   10mM Tris.CL,pH7.6
   10mM $MgCl_2$

2.14.2 Bacterial Culture Media

1. L-broth.

   1% tryptone
   0.5% yeast extract
   0.5% NaCl
   0.1% glucose

2. L-agar.

   1% tryptone
   0.5% yeast extract
   0.5% NaCl
   1.5% agar

3. Top agarose

   1% tryptone
   0.5% NaCl
   0.7% agarose

4. 2xTY medium

tryptone	16g/l
yeast extract	10g/l
NaCl	5g/l

5. H-Agar.

tryptone	10g/l
NaCl	8g/l
agar	12g/l

6. H-Top agar. as above except only 8g/l of agar.

7. Ampicillin

   Made up as 100x stock (5mg/ml) in ampicillin solvent and stored for up to 1 week.

8. Ampicillin solvent

   19:1 mixture of 0.1M Na_2HPO_4 and 0.1M NaH_2PO_4 filter sterilised (Millipore 22µm filters).

2.14.3 Genomic and cDNA Libraries

The genomic clone λP1 was isolated from a HaeIII/ SauI partial genomic library cloned into Charon 4A (Lawn et al.1978). DNA was ligated to EcoRI linkers for cloning.

pPGDH1 was isolated from a λgt10 cDNA library made from RNA from normal adult lung. Inserts were ligated to vector by means of EcoRI linkers. This library was a kind gift from B.Wainwright (St.Mary's Hospital) for whom the library was made (by Transgene).

pPGDH2 was isolated from a λgt11 cDNA library made from upper duodenum RNA. This was kindly provided by Y.Edwards (MRC Human Biochemical Genetics Unit, Galton Laboratory). This library was also constructed with EcoRI linkers.

λTC12 was picked from a λgt10 library made with RNA from a T cell line (MOLT17). This was constructed by T.Rabbitts (LMB, Cambridge) again with the use of EcoRI linkers.

Genotypes of Bacterial Strains

JM101	*sup*E*thi*Δ(lac-proAB) F'[*tra*D*pro*AB^{+}*lac*I^{q}*lac*Z M15]
LE392	*sup*E44*sup*F58*hsd*R514*gal*K*gal*T22 *met*B1*trp*R55*lac*Y1
NM514	*hsd*Δ5Δ(*lac-pro*) F'*lac*ZM15*lac*I^{q}

CHAPTER 3 RESULTS

3.1 Rat PGD cDNA

The starting point for the isolation of human PGD sequences was an 880bp partial cDNA for rat PGD (PGD-1) isolated by Miksicek and Towle (1983). This was isolated from a rat hepatic cDNA library constructed from RNA from rats fed on a high carbohydrate, fat-free diet and administered triiodothyronine (T_3). These conditions induce lipogenic enzymes such as PGD, primarily by increasing mRNA synthesis (Miksicek & Towle,1982). A single clone was isolated on the basis of differential hybridisation to PGD-enriched and total hepatic RNA. All the steps taken in the selection of the clone, pPGD-1, were based solely on an antibody specific for PGD which was isolated from the sera of rabbits repeatedly challenged with purified rat PGD. Incomplete knowledge of the specificity of the antibody allows for some ambiguity in the identity of pPGD-1. This was a consideration in the subsequent characterization of human sequences isolated on the basis of homology to pPGD-1.

Miksicek and Towle (1983) showed by Northern blot analysis that pPGD-1 detected a 2.4kb polyadenylated transcript which displayed a tissue distribution in rat which was in accordance with that for PGD enzyme activity. Although this represented indirect evidence

that the pPGD-1 clone genuinely represents a 6-PGD cDNA, other NADPH-dependent dehydrogenases such as G6PD and malic enzyme have similar tissue distributions (Glock & McLean).

Another potential problem in the identification was the small proportion of PGD mRNA which pPGD-1 contains. The rat PGD enzyme is known to be a homologous dimer made up of 52 kDalton subunits (Bridge et al.1975). A rough calculation suggests that these must comprise of about 490 amino acids and that approximately 1500 nucleotides of the mRNA are required to encode them. Thus, given a 2.4kb mRNA, a substantial proportion, if not all, of pPGD-1 may be made up of untranslated sequence. In general, the non-coding sequences at the 3' end of the mRNA are much longer than the 5' leader sequences so that, assuming that the cDNA was primed from the region of the poly(A) tail, pPGD-1 might be entirely composed of 3' non-coding sequences. It seems reasonable to assume that the selective constraint on such non-coding sequence is less stringent than that acting upon coding sequence. In that case, it follows that the degree of homology between pPGD-1 and its human counterpart may be insufficient to permit the use of pPGD-1 as a heterologous probe.

3.2 Isolation of Human PGD Sequences

The initial characterization of pPGD-1 did indeed suggest that it includes non-coding DNA:

1. Hybridisation of pPGD-1 to human genomic DNA revealed only weakly homologous sequences although more intense signals were seen in hamster DNA (see later).
2. Two total genomic libraries (4 genome equivalents) and a chromosome 1-sorted library (32 chrom.1 equivalents) were probed with pPGD-1 at moderate stringency (2xSSC/65°C). None of the clones screened hybridized.

The isolation of rat mRNA sequences which lie further 5' of pPGD-1 by primer extension was attempted.Double-stranded pPGD-1 insert was denatured and annealled at 50°C to rat hepatic poly(A)$^+$ RNA in 80% formamide which favours the formation of RNA:DNA hybrids. In principle, once the cDNA has annealed to the PGD mRNA, its 3' end can be extended using the mRNA as template. Thus the nucleic acids were precipitated and incubated with with reverse transcriptase in the presence of radioactive dCTP as described by Favoloro et al.1981. However, when the products of this reaction were hybridized back to a rat genomic clone thought to contain PGD coding sequences, no PGD-specific extension was detected. There was some doubt, however over the authenticity of the genomic clone.

In the event, despite the weak homology between pPGD-1 and human genomic sequences, pPGD-1 hybridized

to a single independent clone (out of 10^6 screened) in a human lung cDNA library (Transgene, see Materials). This single phage was isolated, its 0.65kb insert was removed via its artificial EcoRI ends and subcloned into pAT153. This plasmid was named pPGDH1 (H1).

Subsequently H1 was used to probe 10^6 phage of an amplified human upper duodenum cDNA library (Dr.Edwards; see Materials) at high stringency (washed in 0.1xSSC/65°C) which had previously been unsuccessfully screened with the rat cDNA. H1 hybridized to approximately 200 clones (consistent with the estimated abundance of PGD mRNA of about 0.015%- Miksicek & Towle, 1982). Twelve of these were isolated and analyzed. All twelve were found to contain identically-sized inserts of 0.9kb. One such insert was cloned into the EcoRI site of pAT153 (pPGDH2).

H2 was then used to screen an amplified T cell cDNA library (Dr.Rabbitts; see Materials) at high stringency. It hybridized to about 100 clones (out of 10^6 screened), twelve of which were isolated and analyzed. All but two contain inserts of 0.9kb or less. λTC11 contained a 1.15kb insert and λTC12 contained a 1.7kb insert. Both possessed a single internal EcoRI site such that when λTC12 was digested with EcoRI in order to separate vector from cDNA, two cDNA fragments of 0.8 and 0.9kb are produced. These were subcloned separately into pAT153 and called pPGDH3 (0.8kb) and pPGDH4 (0.9kb). Restriction maps of the four subcloned cDNAs (pPGDH1-4) were constructed and aligned. The composite map of all four cDNAs is shown in figure 1.

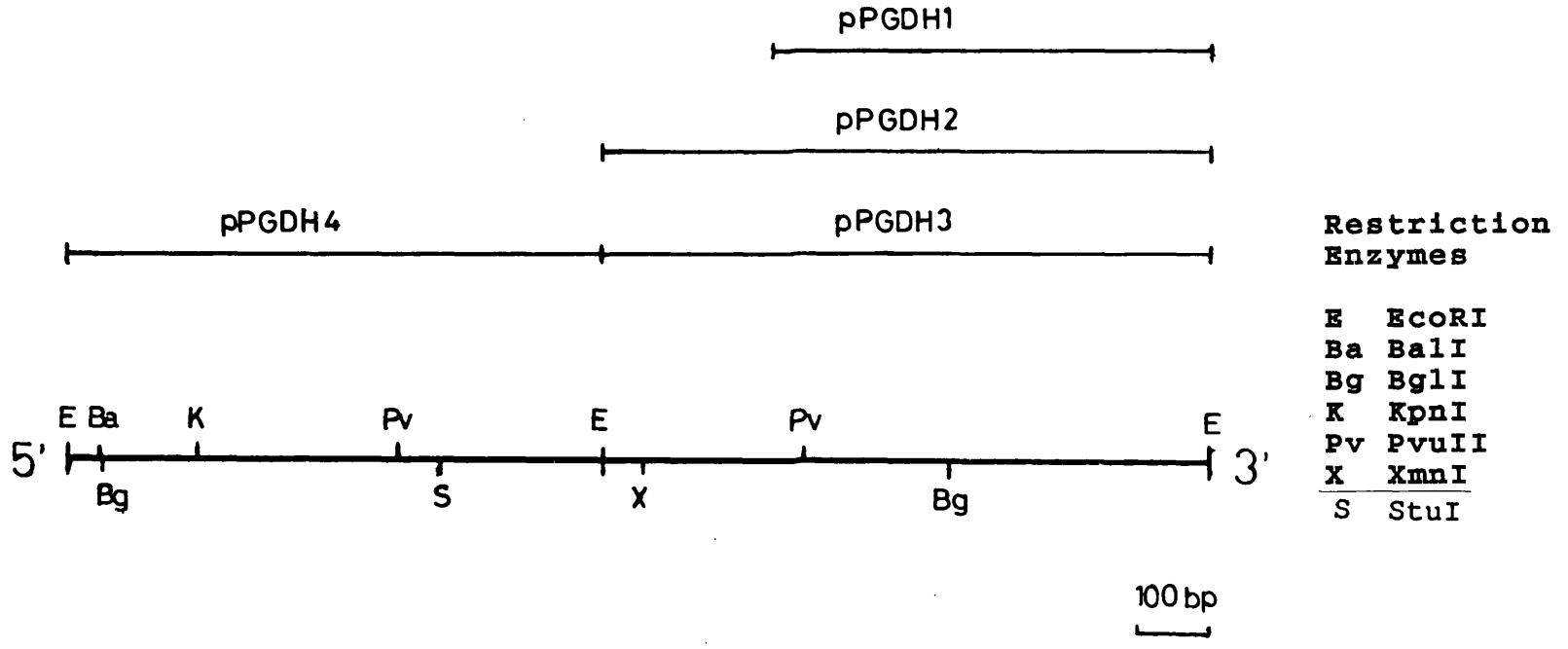


Figure 1. Restriction Map of Human PGD cDNAs.

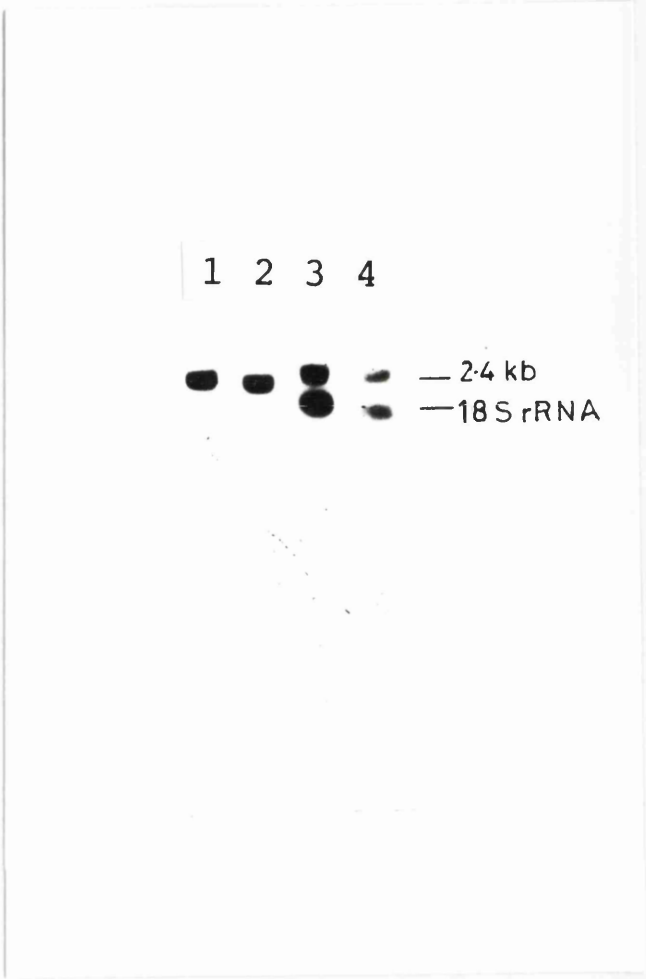
The length of the longest cDNA isolated (λTC12) is 1.7kb. The 5' 800bp EcoRI fragment (pPGDH4) and the 3' 900bp EcoRI fragment (pPGDH3) were subcloned separately. pPGDH1 and pPGDH2 are independently cloned cDNAs. Sequence and restriction map comparisons showed that H2 and H3 are identical and that H1 is a shortened copy of both as indicated above. The direction of transcription was determined by sequencing of the open reading frame in H4 and the 5' end of H3 (see section 3.5).

The restriction maps of H2 and H3 were found to be identical. H1 hybridizes strongly to H2 and H3 and restriction mapping indicated substantial homology with H2/H3. Sequence information (see later) confirmed that H2 and H3 are identical and that H1 is a shorter copy of both as illustrated in figure 1.

No open reading frame was found in H1, the cDNA detected by the rat probe, pPGD-1. This supports the initial evidence suggesting that the rat cDNA represents 3' untranslated sequence. However, the lack of a polyA tail raised the possibility that this was not the case. The finding that H2 and H3 are identical, even though they were isolated from different libraries, sheds some light on this apparent discrepancy. Both these cDNAs were cloned into gt10 using EcoRI linkers. Their exact correspondence probably reflects the existence of two EcoRI sites in PGD mRNA separated by 0.9kb. Assuming incomplete protection by methylation of internal EcoRI sites, H2 and H3 may be natural, internal EcoRI fragments created after EcoRI digestion to remove excess linkers. Thus, it is certainly feasible that the original mRNA possesses a polyA tail but that it lies beyond one of the EcoRI sites.

3.3 Northern Blot Analysis of RNA

Figure 2 shows the hybridisation of H4 to Northern blots of total and poly(A)$^+$ RNA from different cell lines. In lanes 3 and 4 two bands are clearly visible. The lower one in each case comigrates with the 18S ribosomal RNA band. This band is the result of non-specific hybridisation of the probe to 18S rRNA. This is clearly illustrated by the detection of a single band in poly(A)$^+$ RNA (lanes 1 and 2). Using the 18S rRNA as a marker (which has been found to co-migrate with mRNA of 2.1kb; BRL Focus vol.8(1)p.9), the band above can be sized at 2.4kb, the same size as rat PGD mRNA. This concords with the similar molecular weights of the rat and human PGD subunits (52kDal; Bridge et al.)



Lanes:

1. **EJ Poly(A)$^+$ RNA**
2. **MGL8B2 Poly(A)$^+$ RNA**
3. **EJ total RNA**
4. **MGL8B2 total RNA**

Figure 2. Hybridisation of pPGDH4 to Human RNA

The hybridisation of the 5' human cDNA clone, pPGDH4 to Northern blots of human poly(A)$^+$ and total RNA is shown. EJ is a human bladder carcinoma cell line and MGL8B2 is a human lymphoblast cell line. H4 detects a single 2.4kb band in poly(A)$^+$ RNA in both cell lines. The second smaller band seen in total RNA represents the non-specific hybridisation of H4 to 18S ribosomal RNA. 18S rRNA comigrates with mRNA of 2.1kb.
Visualisation of the rRNA on the filter under U.V showed that the apparent size differences between lanes were the result of gel running artifacts.

3.4 Hybridisation of cDNAs to Genomic DNA

Figure 3 compares the genomic hybridisation patterns of the rat and human cDNAs which show homology (pPGD-1 and H1).

Figure 3a (lanes 3-5) demonstrate that pPGD-1 weakly detects human genomic sequences at low stringency (2xSSC/65° C) and as expected, H1 detects the same bands (Fig.3b, lanes 2-4) at higher stringency (0.1xSSC/65° C). Both detect 3.5kb, 1.25kb and 1.0kb EcoRI bands.

The discovery that both cDNAs detect 1.25kb EcoRI fragments in human and rat DNA (fig.3a, lanes 4-6 and fig.3b, lanes 2-5; Miksicek & Towle (1983) showed that the rat cDNA detects a 1.25kb EcoRI fragment in rat genomic DNA) initially indicated that the there may be a region of exact homology between the two species at least 1.25 kb long.

The fact that H1 detects three EcoRI bands in genomic DNA despite not possessing any internal EcoRI sites itself reflects two possibilities:

1. H1 may contain sequences (exons) which are interrupted (by introns) in the genome, although the evidence presented earlier in favour of H1 being non-coding seems to contradict this.
2. H1 detects several unlinked loci. The most likely situation in which this would occur would be if PGD pseudogenes existed. i.e. sequences closely related to

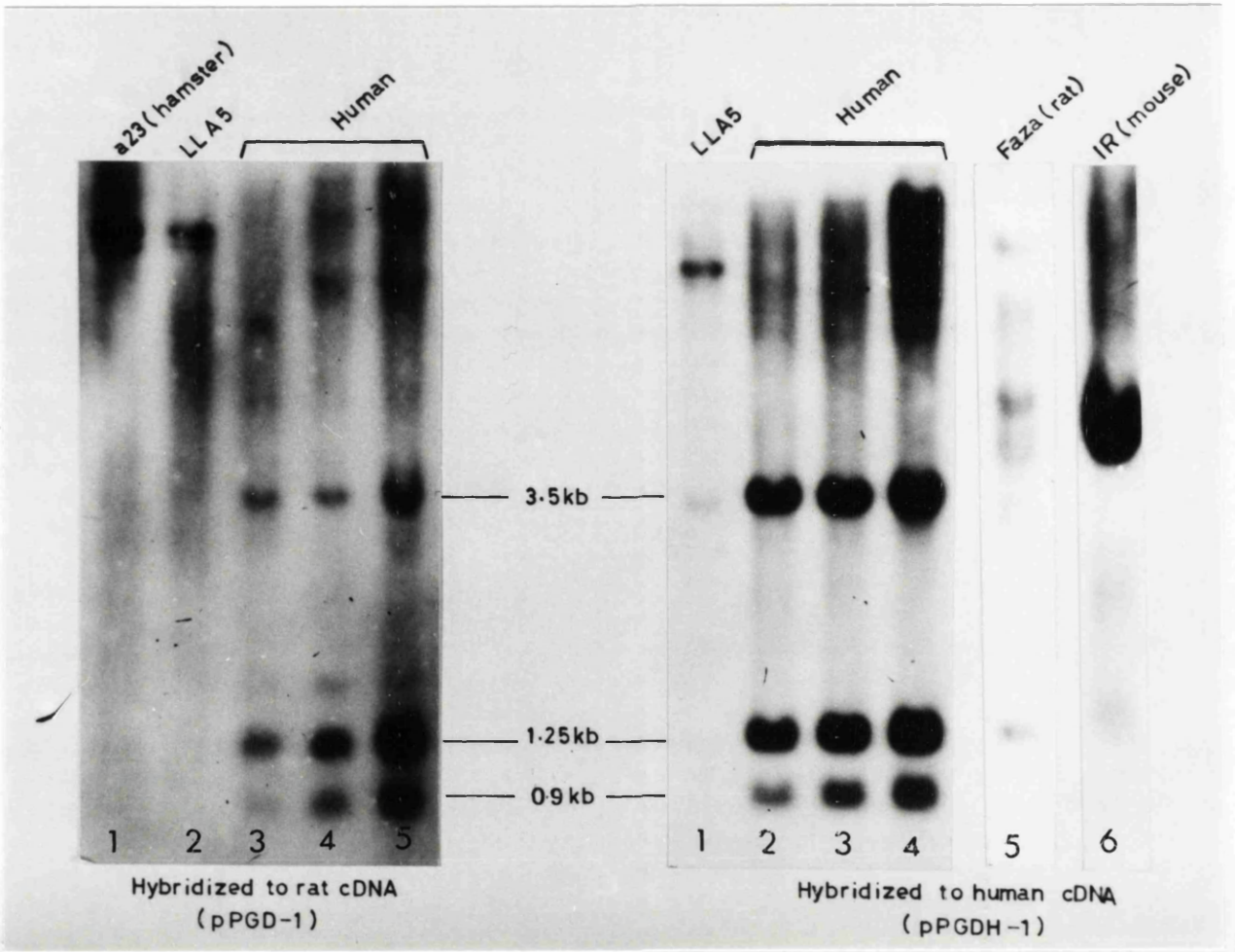


Figure 3. Comparison of hybridisation of rat and human cDNAs to genomic DNA.

the PGD gene, but not expressed. These have been reported for a number of 'household' enzymes (see Vanin,1985 for review).

In order to distinguish between these two possibilities several approaches can be taken. Initially, human genomic DNA was digested with several different restriction enzymes and probed with H1. Figure 4 shows the results of this analysis. All lanes show at least three bands and PvuII which produces four H1-hybridizing fragments also cuts the probe once. This strongly supports the hypothesis that H1 detects several unlinked loci. This was confirmed by somatic cell hybrid analysis.

Lanes: 1. HindIII.
2. PvuII.
3. PstI.
4. BamHI.
5 & 6. EcoRI.

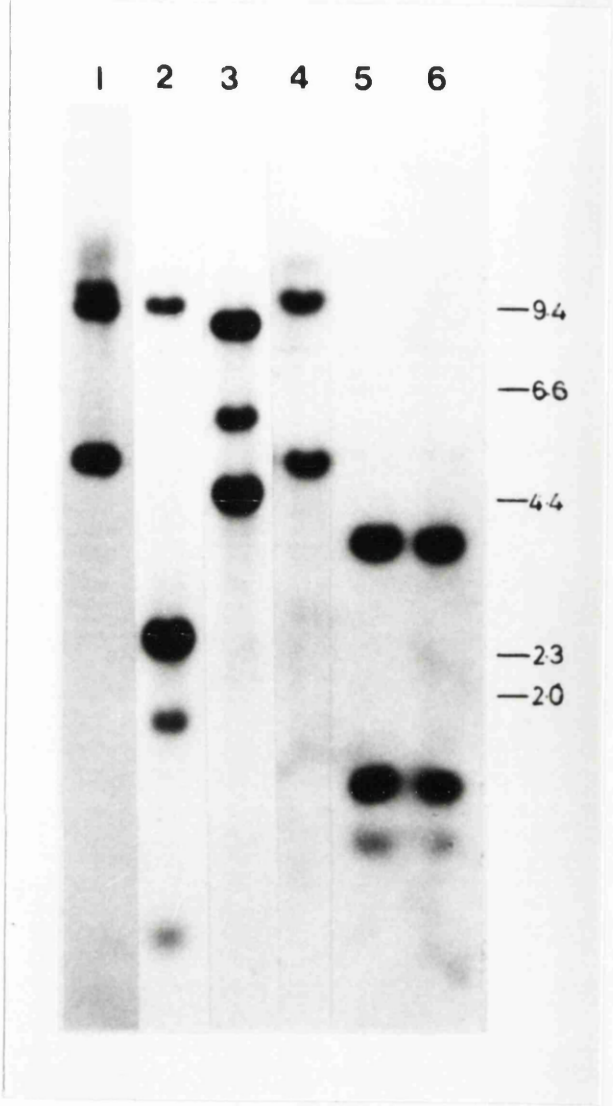


Figure 4. Human Genomic DNA Digests vs. pPGDH1.

This shows that the human 3' cDNA clone, pPGDH1, detects at least three fragments in genomic DNA digested with various restriction enzymes. All but one of these restriction enzymes do not cut within H1 and generate three H1-hybridizing fragments. H1 detects four fragments in DNA digested with PvuII which recognizes one site in H1. All of these hybridisations were washed to a final stringency of 0.1xSSC/65° C.

3.5 Assignment of Loci detected by pPGDH1

3.5.1 Somatic Cell Hybrids

Figure 5 shows an example of hybridisation of H1 to EcoRI-cut DNA from somatic cell hybrids. Lane 1 shows the hybridisation pattern in total human genomic DNA. The parent hamster line, a23 (lane 2) of the hybrids in lanes 3-8 is shown for comparison.

These hybrids clearly show the independent segregation of the three EcoRI bands and thus confirm that H1 detects several loci on different chromosomes.

Table 1 summarizes the results of the analysis of 18 somatic cell hybrids showing their human chromosome content and their hybridisation pattern when digested with EcoRI and probed with H1.

In all hybrids, the 3.5kb EcoRI band segregates unambiguously with chromosome 1, whilst the 1.25kb band segregates with chromosome 18. On the other hand, the 1.0kb band shows at least some discordancy with every chromosome. This may be due the presence of several such fragments on different chromosomes. However, no combination of two chromosomes satisfies these data either. The assignment of this locus (or loci) was not pursued as it clearly did not segregate with chromosome 1 and its precise location was considered to be of low priority in the context of this project.

The Taxi and HCH hybrids carry translocations involving human chromosome 1. The breakpoints in the

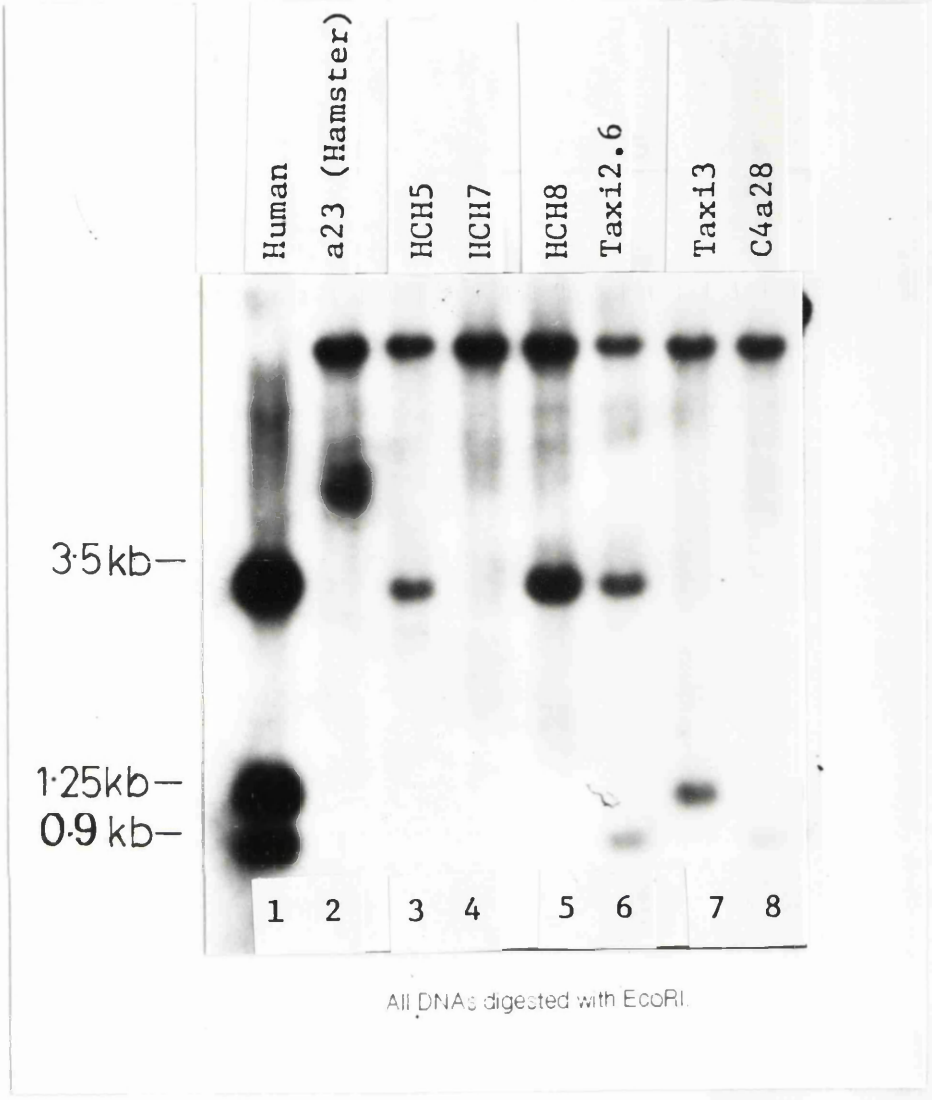


Fig.5. Hybridisation of H1 to a Panel of Somatic Cell Hybrids.

All DNAs were digested with EcoRI.
H1 detects 3 bands in EcoRI-cut human genomic DNA (lane 1). A single, high molecular weight band is consistently seen in a23 (lane 2) and the hamster-human hybrids (lanes 3-8). The three human loci detected by H1 clearly segregate independently in the hybrids. The other bands visible in all lanes represent partially-digested fragments.

Table 1. Hybrid Panel vs. pPGDH1 cDNA.

The human cDNA, pPGDH1, was hybridized to EcoRI digests of DNA extracted from 18 different somatic cell hybrid lines. The table shows the human chromosome complement of each hybrid. The final three columns indicate the human EcoRI fragments detected by H1 in each hybrid. The 3.5kb and 1.25kb EcoRI bands unambiguously segregate with chromosomes 1p36.13-1pter and 18 respectively. The 0.9kb band, on the other hand, cannot be assigned to any single chromosome.

A '/' indicates that the hybrid was not typed for that particular chromosome. The HCH and Taxi hybrids were made with human donors that possessed translocations involving the distal region of chromosome 1p. The letters A-D refer to the fragment of chromosome 1 contained in these hybrids as indicated below the table.

CHROMOSOMES / HYBRIDS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	EcoRI bands detected by H1: 3.5 kb	1.25 kb	0.9 kb
DIS 16	-	-	-	-	-	-	+	-	/	-	+	+	-	+	-	+	+	+	-	-	+	+	-	-	+	-
DIS 9	-	-	-	-	/	-	-	+	-	-	-	-	-	-	-	/	+	-	+	-	-	-	-	-	-	-
B4-2	+	-	+	+	+	-	+	+	+	+	+	+	+	+	-	-	+	-	-	+	+	+	-	+	-	+
DIS 20	-	-	+	-	+	+	-	-	-	+	+	+	-	-	-	-	+	+	-	-	+	+	-	-	+	+
C4a11	-	-	+	-	-	-	-	+	+	+	-	-	+	+	+	+	+	-	-	-	-	+	-	-	-	-
C4a28	-	-	+	-	-	-	-	+	+	-	+	+	+	+	-	+	+	-	-	-	-	+	+	-	-	+
RVL13	+	+	-	/	/	-	+	+	-	/	-	+	/	+	+	+	+	/	-	+	+	/	+	+	+	+
HCH 5	A	-	-	-	/	/	+	-	-	-	-	+	/	+	-	/	+	-	/	-	/	/	-	+	-	-
HCH 7	B	-	-	+	/	/	+	/	-	/	-	+	/	-	-	/	+	-	/	-	/	/	-	-	-	-
HCH 8	A	-	-	/	/	/	-	-	-	/	-	-	/	+	-	/	+	-	/	-	/	/	-	+	-	-
Taxi 2.6	C	+	+	+	/	+	+	+	+	+	-	+	/	+	+	/	+	-	+	+	/	/	+	+	-	+
Taxi 3	D	-	-	/	/	+	+	+	-	+	-	+	/	-	-	/	+	+	+	-	/	/	-	-	+	-
Taxi 8	D	-	+	+	/	-	-	-	-	-	-	-	/	+	-	/	+	-	+	-	/	/	-	-	-	+
HORL4.11.B6P	+	-	+	-	-	-	-	-	-	+	+	-	+	-	+	-	-	+	-	-	+	+	+	+	+	-
HORL4.11.B5N	-	-	+	-	-	-	-	-	-	+	+	-	+	-	+	-	-	+	-	-	+	+	+	+	-	-
CTP34B4	+	+	+	-	+	+	+	+	-	-	-	+	-	+	-	+	+	+	-	-	-	-	+	+	+	-
FHA7	-	-	+	+	+	+	+	+	-	/	-	+	+	/	+	+	+	+	+	+	-	+	+	-	+	+
C4a26	-	-	+	-	-	-	-	+	+	-	/	/	+	+	+	-	+	-	-	-	-	+	-	-	-	+

A contains 1pter-1p36.13 translocation chromosome.
B contains 1p36.13-1qter translocation chromosome.
C contains 1pter-1p34 translocation chromosome.
D contains 1p34-1qter translocation chromosome.

two series of hybrids have been assigned cytologically to 1p34 and 1p36.13 respectively (Carritt et al,1982). Thus, once the assignment of the 3.5kb band to chromosome 1 had been made this was refined by hybridisation of H1 to these hybrids (see Figure 5 and Table 1). H1 detects a 3.5kb EcoRI band in HCH5 and HCH8 which possess the 1p36.13-1pter translocation chromosome but fails to in HCH7 which contains 1qter-1p36.13. This improves the assignment of the 3.5kb EcoRI band to 1p36.13-1pter. The hybrids used were the same as those used for the most accurate assignment of PGD to date (Carritt et al.1982).

The hybrid DIS 2.6 was made with a karyotypically normal human cell line, but no human chromosomes are visible cytologically. However, it is known (Carritt et al.1982) that it does contain a fragment of chromosome 1, including the PGD locus. This hybrid has subsequently been tested for the presence of a total of 10 other 1p markers (Carritt et al.1982; N.Dracopoli, personal comm.). None of these are present in DIS 2.6:

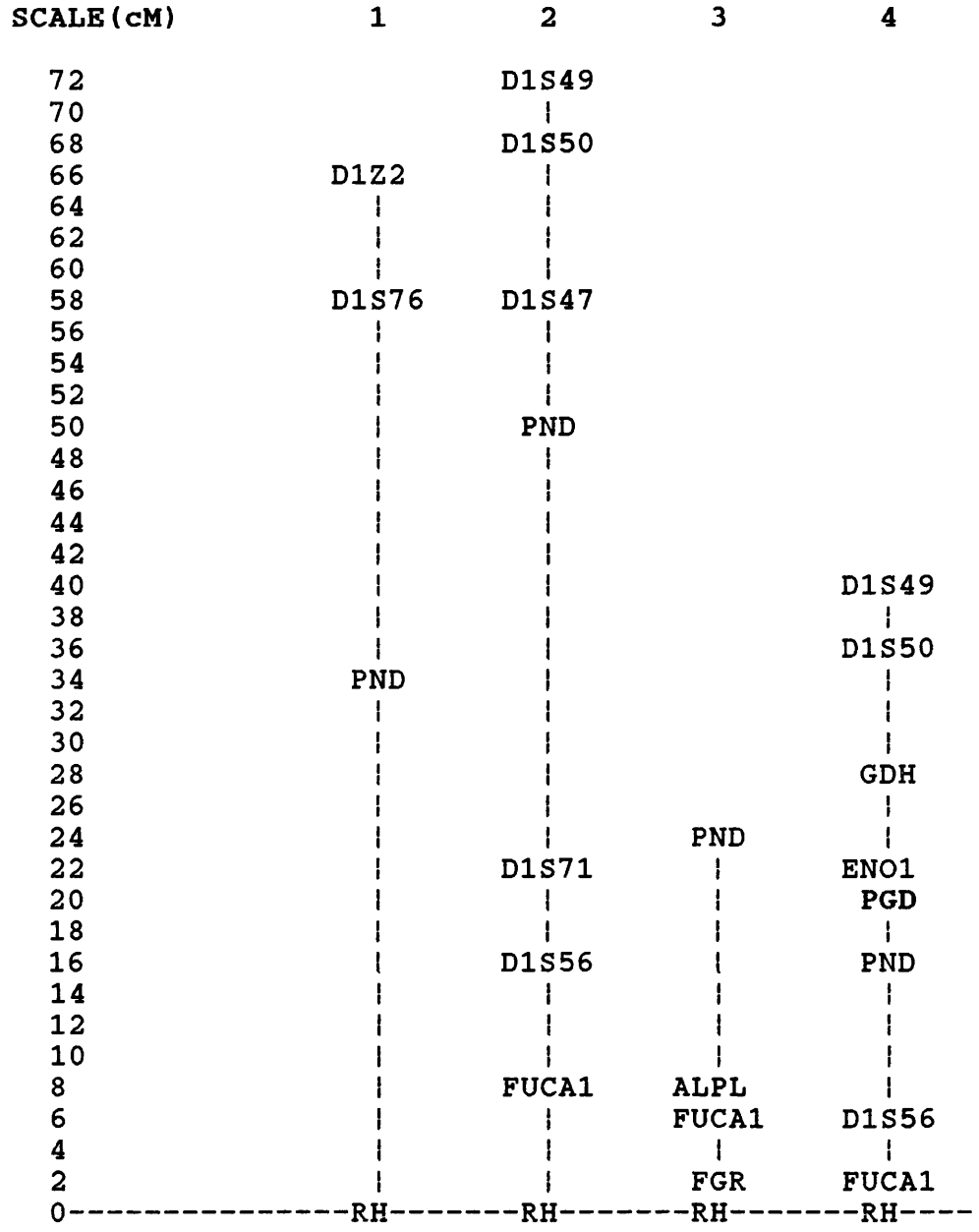
pter	DIS 2.6
¦	
GDH	-
D1S47	-
ENO1	-
PGD	+
D1Z2	-
PND	-
ALPL	-
FUCA1	-
FGR	-
MYCL1	-
GLUT1	-
¦	
cen	

See figure 6 for the most recent genetic maps of chromosome 1p including most of these markers.

Note that FUCA1, originally scored as present in DIS2 by enzyme analysis (Carritt et al.1982), is here scored as absent in the subclone DIS2.6 by Southern blot analysis using an α-fucosidase cDNA. Whether this reflects a genuine difference between the DIS2 and its subclone or an unreliable enzyme typing is not certain.

The present status of the genetic map of chromosome 1 (see figure 6; Bruns & Sherman,1989; Keats et al.,1989) gives little information concerning the genetic distances between these markers. However a genetic distance of 6cM (based on pairwise and multipoint data) has been estimated between ENO1 and PND. This suggests that DIS 2.6 contains, at most, a chromosome fragment corresponding to 6cM. In 1p36 this is likely to be less than $6x10^6$ bp of DNA (see Discussion). H1 detects only the 3.5kb EcoRI fragment in DIS 2.6.

Thus, the assignment of this fragment to a small stretch of DNA within 1p36.13-1pter to which PGD has previously been assigned and the fact that H1 shares homology with a rat PGD cDNA suggests that, with little doubt, H1 represents part of the PGD mRNA.



1.O'Connell et al.(1989)
2.Donis-Keller et al.(1989)
3.Dracopoli et al.(1988)
4.Keats et al.(1989)

Figure 6. Genetic Maps of The Distal Region of Chromosome 1

Maps 1-3 are sex-averaged and were constructed by multipoint analysis assuming no interference. No.4 is a male map based on pairwise data as well as information on order obtained from multipoint linkage analysis. All maps are consistent with respect to the order of shared markers but estimates of distance vary greatly. The order of ENO1, PGD and PND and the distances separating them are illustrated in map 4.

3.5.2 In Situ Hybridisation

The assignment of a H1-hybridizing locus to 1p36 was confirmed by *in situ* hybridisation of tritiated pPGDH1 to lipsol-banded metaphase chromosomes. This was carried out by K.Gulati (Galton Laboratory) as described in Methods.

66 complete metaphase spreads were analyzed. Figure 7 shows an example of a banded metaphase spread prior to hybridisation and the same chromosomes after hybridisation and exposure. Figure 8 is an ideogram of the 346 grains located. Of these, 6.7% were within 1p35-1p36, 4.1% were within 18q12-22 and 3.8% were within 6p11-6q11. No other significant peaks were detected although the generally high background may be related to the existence of other loci with homology to H1.

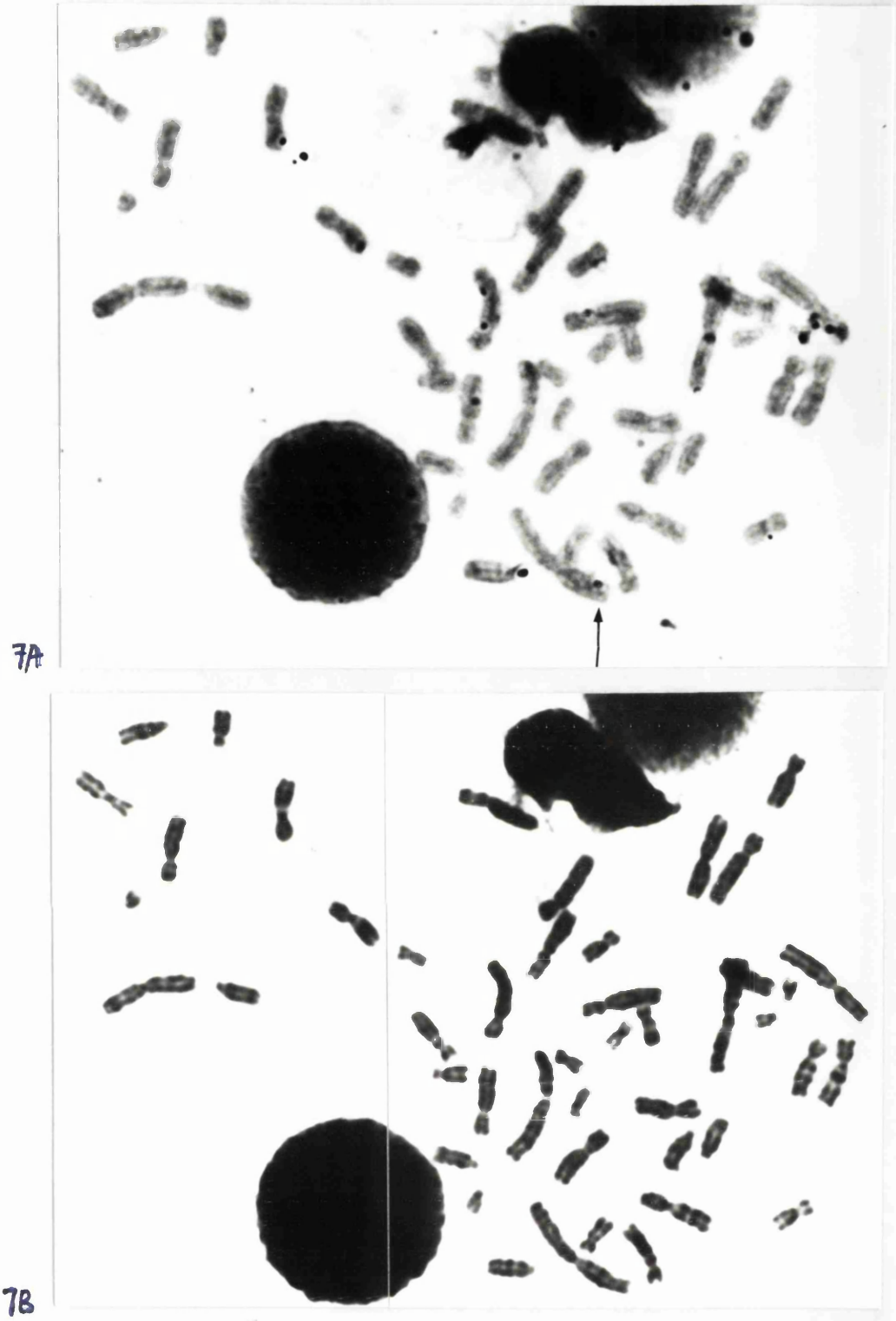


Figure 7. *In Situ* Hybridisation of Tritiated pPGDH1 to Metaphase Chromosomes

Fig.7A shows a metaphase spread which has been hybridized to tritiated pPGDH1. The arrow indicates a grain on 1p36. Seven other grains on this spread were scored.

Fig.7B shows the same metaphase lipsol-banded prior to hybridisation.

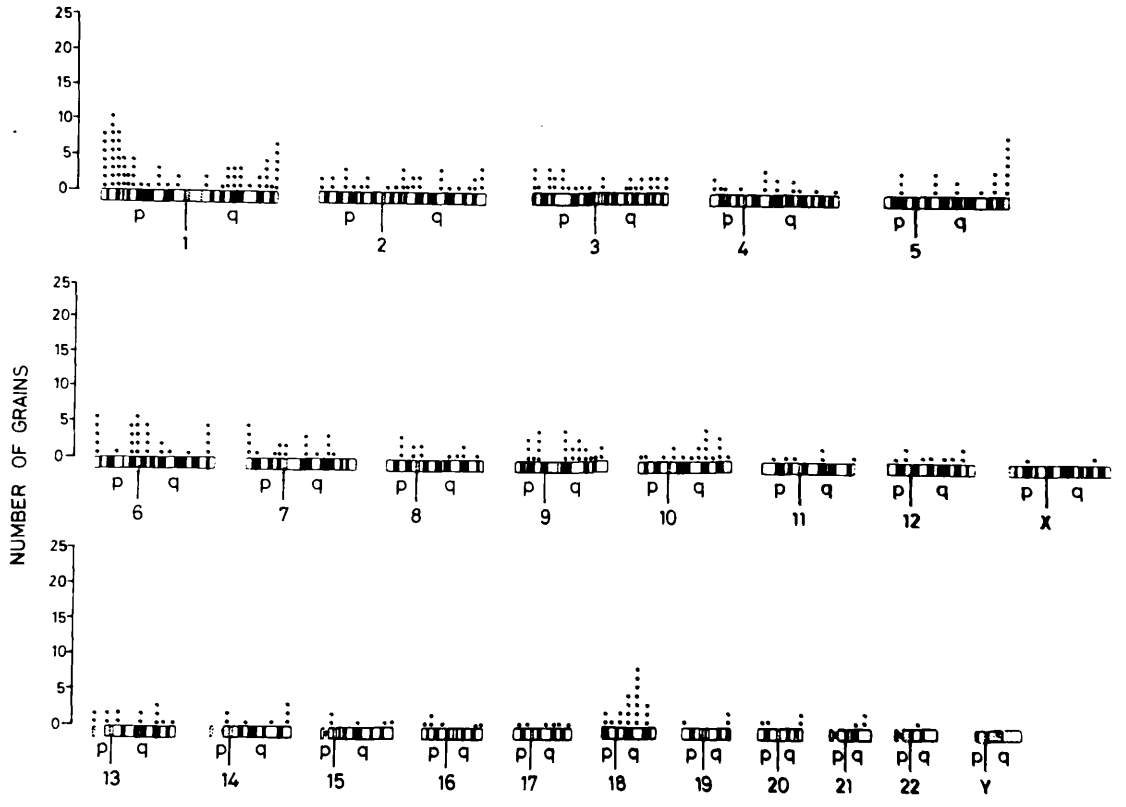


Figure 8. Ideogram of the Results of the *In Situ* Hydridisation of Tritium-Labelled pPGDH1 to Metaphase Chromosomes

66 Metaphases were analyzed and 345 grains in contact with chromosomes were scored (average= 5 grains/cell). Of these, 6.7% were within 1p35-p36. Minor peaks at 18q12-q22 (4.1%) and 6p11-6q11 (3.8%) were also observed. The generally high level of background may reflect the presence of other loci homologous to H1.

3.6 Assignment of Loci Detected by pPGDH4

3.6.1. Somatic Cell Hybrids

Since H4 does not hybridize to H1, there was a possibility that H4 was unrelated to H3 and that their co-presence in λTC12 was a cloning artifact (see figure 1). Therefore H4 was hydridized to a panel of hybrids in order to determine whether it detected a locus in 1p36.

Figure 9 shows total human genomic DNA (lane 1), a panel of hybrids (lanes 2-8) and a23, the parent hamster cell line of the hybrids (lane 9) digested with EcoRI and hybridized to H4. H4 detects 5 human fragments of 10.5, 6, 4.5, 3 and 0.75kb. The three biggest bands segregate with chromosome 1 as is shown in Table 2. The 3kb and 0.75kb bands were not assigned unambiguously.

Since H4 detects three EcoRI fragments at the same locus and yet does not possess internal EcoRI sites, this strongly suggests that H4 contains coding sequences which are interrupted in the genome. As this is not the case for H1, the most likely explanation is that H4 is 5' of H1 (and H2/H3) such that the the orientation of the cDNA is as shown in figure 1. Given an mRNA length of 2.4kb, the maximum distance between the 5' end of H4 and the 5' end of the mRNA is about 700bp. It may be less than this if the 3' ends of H1, H2 and H3 do not correspond to the natural 3' end of the mRNA as was suggested earlier.

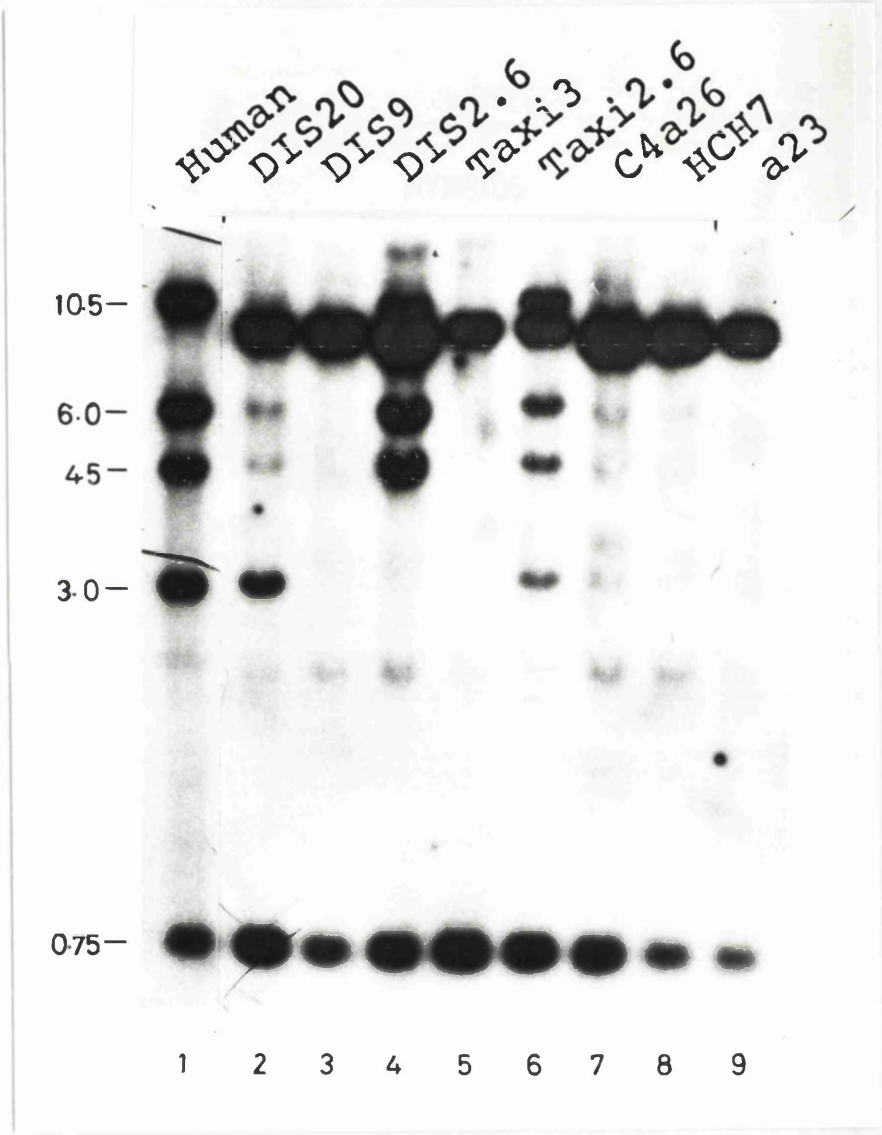


Figure 9. Hybridisation of pPGDH4 to a Panel of Somatic Cell Hybrids.

All DNAs were digested with EcoRI and hybridized to H4 at moderate stringency (2xSSC/65°C final wash).
H4 detects at least five bands in EcoRI-cut human DNA (lane 1). The three largest fragments segregate together independently of the 3.0kb band.

A faint 4.0kb band in lane 7 probably represents a partially-digested fragment.

CHROMOSOMES HYBRIDS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	EcoRI bands 10.5, 6.0 & 4.5kb EcoRI bands detected by H4
DIS 9	-	-	-	-	/	-	-	+	-	-	-	-	-	-	-	/	+	-	+	-	-	-	-	-
B4-2	+	-	+	+	+	-	+	+	+	+	+	+	+	+	-	-	+	-	-	+	+	+	-	+
DIS 20	-	-	+	-	+	+	-	-	-	+	+	+	-	-	-	-	+	+	-	-	+	+	-	-
C4a11	-	-	+	-	-	-	-	+	+	+	-	-	+	+	+	+	+	-	-	-	-	+	-	-
HCH 7	B	-	-	+	/	/	+	/	-	/	-	+	/	-	-	/	+	-	/	-	/	/	-	-
HCH 8	A	-	-	/	/	/	-	-	-	/	-	-	/	+	-	/	+	-	/	-	/	/	-	+
Taxi 2.6	C	+	+	+	/	+	+	+	+	+	-	+	/	+	+	/	+	-	+	+	/	/	+	+
Taxi 3	D	-	-	/	/	+	+	+	-	+	-	+	/	-	-	/	+	+	+	-	/	/	-	-
CTP34B4	+	+	+	-	+	+	+	+	-	-	-	+	-	+	-	+	+	+	-	-	-	-	+	+
C4a26	-	-	+	-	-	-	-	+	+	-	/	/	+	+	+	-	+	-	-	-	-	+	-	-

A contains 1pter-1p36.13 translocation chromosome.
B contains 1p36.13-1qter translocation chromosome.
C contains 1pter-1p34 translocation chromosome.
D contains 1p34-1qter translocation chromosome.

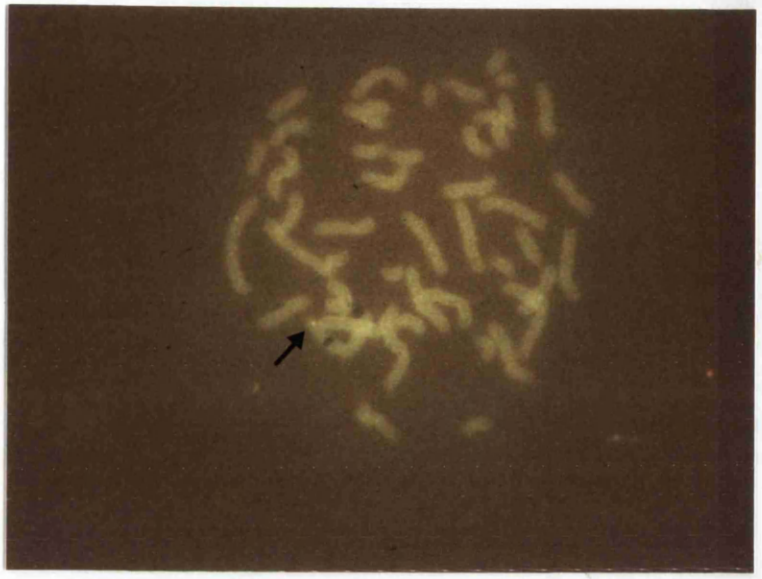
Table 2. Hybrid Panel vs. pPGDH4 cDNA.

The human cDNA, pPGDH4, was hybridized to EcoRI digests of DNA extracted from 10 different somatic cell hybrid lines. The table shows the human chromosome complement of each hybrid. The final column indicates whether the hybrid contained the H4-hybridizing 10.5, 6.0 and 4.5kb human EcoRI fragments. These three fragments unambiguously segregate with 1p36.13-1pter. A '/' indicates that the hybrid was not typed for that particular chromosome. The HCH and Taxi hybrids were made with human donors that possessed translocations involving the distal region of chromosome 1p. The letters A-D refer to the fragment of chromosome 1 contained in these hybrids as indicated below the table.

3.6.2 In Situ Hybridisation

The assignment of an H4-hybridizing locus to 1p36 was confirmed by *in situ* hybridisation of biotinylated pPGDH4 to lipsol-banded metaphase chromosomes. This was carried out by D.Griffen (Galton Laboratory) as described in Methods.

90 fluorescent signals were scored in 45 cells analyzed and of these, 16 (17.8%) hybridized to 1p36 as shown in Figure 11. No secondary peaks were found on chromosome 18 or any other chromosome. Figure 10 shows an example of a banded metaphase spread prior to hybridisation and again under green light after hybridisation in order to show the fluorescent probe on 1p36. This latter photo was underexposed and the signal is much weaker than it originally appeared.



10A

10B

Figure 10. *In Situ* Hybridisation of Biotinylated pPGDH4 to Metaphase Chromosomes.

Fig.10A shows a metaphase spread which has been hybridized to biotinylated pPGDH4 viewed through a DAPI filter. The arrow indicates a fluorescent signal on 1p36.

Fig.10B shows the same metaphase G-banded prior to hybridisation.

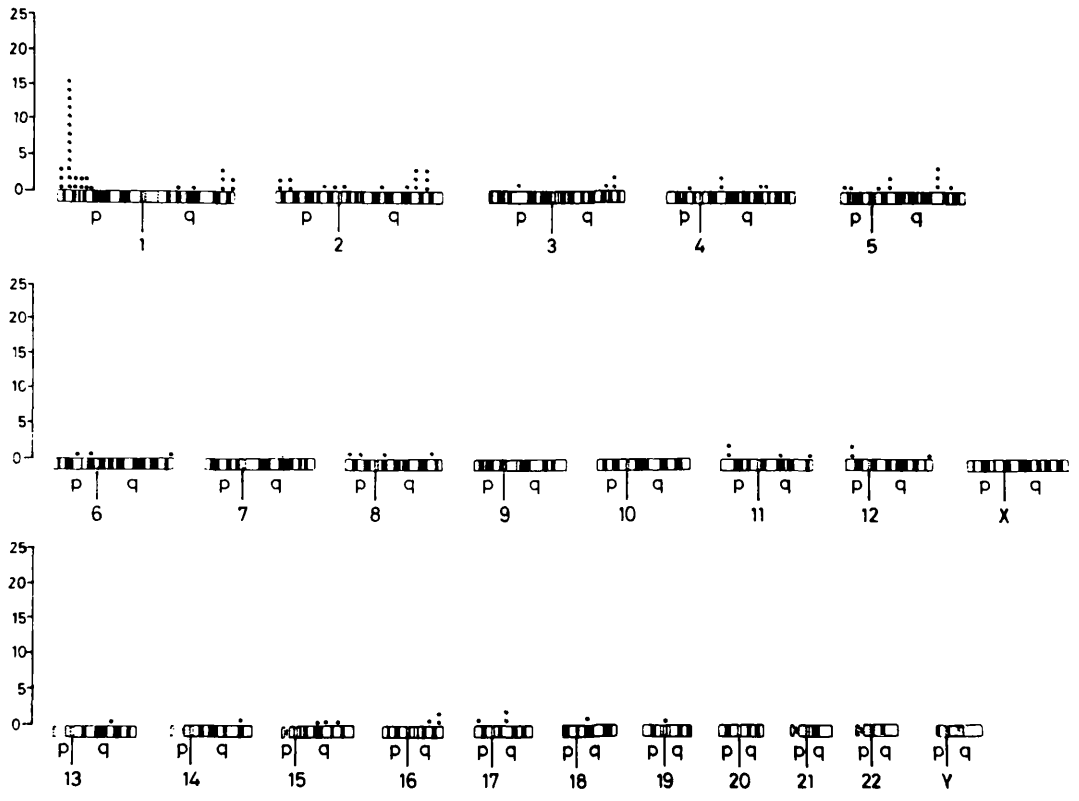


Figure 11. Ideogram of the Results of the *In Situ* Hydridisation of pPGDH4 to Metaphase Chromosomes

90 fluorescent signals were scored in 45 cells analyzed (average=2signals/cell). 17.8% of these were within 1p36. Minor peaks were observed at 2q32-q35 (7.8%) and 1q42-qter (5.6%).

3.7 Detection of an RFLP by pPGDH1

When H1 was isolated, it was used to search for an RFLP. The reasons were twofold. Firstly, the desire to clone PGD initially was based partly on the possibility of generating more genetic information at this locus beyond that detectable at the protein level. Secondly, if an RFLP was found, the establishment of very tight linkage between it and the PGD enzyme polymorphism would provide further evidence for the authenticity of H1.

At least twelve unrelated individuals were screened for RFLPs with twelve different restriction enzymes. Analysis of twelve individuals provides a 72% chance of detecting a minor allele with a frequency of 0.1 ($1-0.9^{12}$) and a chance greater than 90% if the minor allele frequency is more than 0.2 ($1-0.8^{12}$). The enzymes used were MspI, TaqI, EcoRI, BamHI, PstI, BglII, PvuII, RsaI, EcoRV, BstNI, AvaI and XbaI. These enzymes account for about three quarters of all the RFLPs detected to date (Kidd et al.1989).

Of course, these enzymes do not necessarily correspond to the most efficient ones for detecting RFLPs (Devor,1988).

The only one of these restriction enzymes which revealed an RFLP was BamHI. H1 detects 11.0kb and 4.2kb constant bands as well as 11.0 and 9.0kb polymorphic bands in BamHI-digested human DNA. At high stringency the polymorphic 11.0kb band is much more intense than the 11.0kb constant band so that all genotypes are easily scored. This can be seen in fig.12B which shows all

three possible genotypes.

Since H1 detects several unlinked loci, it was obviously necessary to determine which of these is polymorphic for BamHI fragments. Figure 12a shows three hybrids and their parent hamster line, a23, digested with BamHI and hybridized to H1. HCH8 and DIS 2.6 were previously shown to contain only the H1-hybridizing locus on 1p (section 3.5.1). Taxi 2.6 also contains this as well as the unassigned locus. Fig. 12a clearly shows that the fragment of 1p in DIS 2.6 possesses the 11kb BamHI allele whereas the chromosome 1 fragments in HCH8 and Taxi 2.6 both contain the 9.0kb allele. The constant 11kb and 4kb bands cannot be unambiguously assigned because of the co-migration of a23 bands in both cases. Of greater importance, however, is the clear assignment of the RFLP to the PGD locus on 1p36. The RFLP in 11 random individuals is shown for comparison. Figure 13 demonstrates the segregation of the alleles as co-dominant Mendelian markers in a pedigree. The PGD enzyme electrophoretic phenotypes are also shown. The linkage relationship between these two genetic markers is discussed later.

An estimate of allele frequencies in the European population was made from a sample of 54 unrelated individuals: 30 SS, 22 SF and 2 FF. This gives a minor allele frequency of 0.24. The distribution of genotypes correlates well with Hardy-Weinberg predictions for a population in equilibrium.

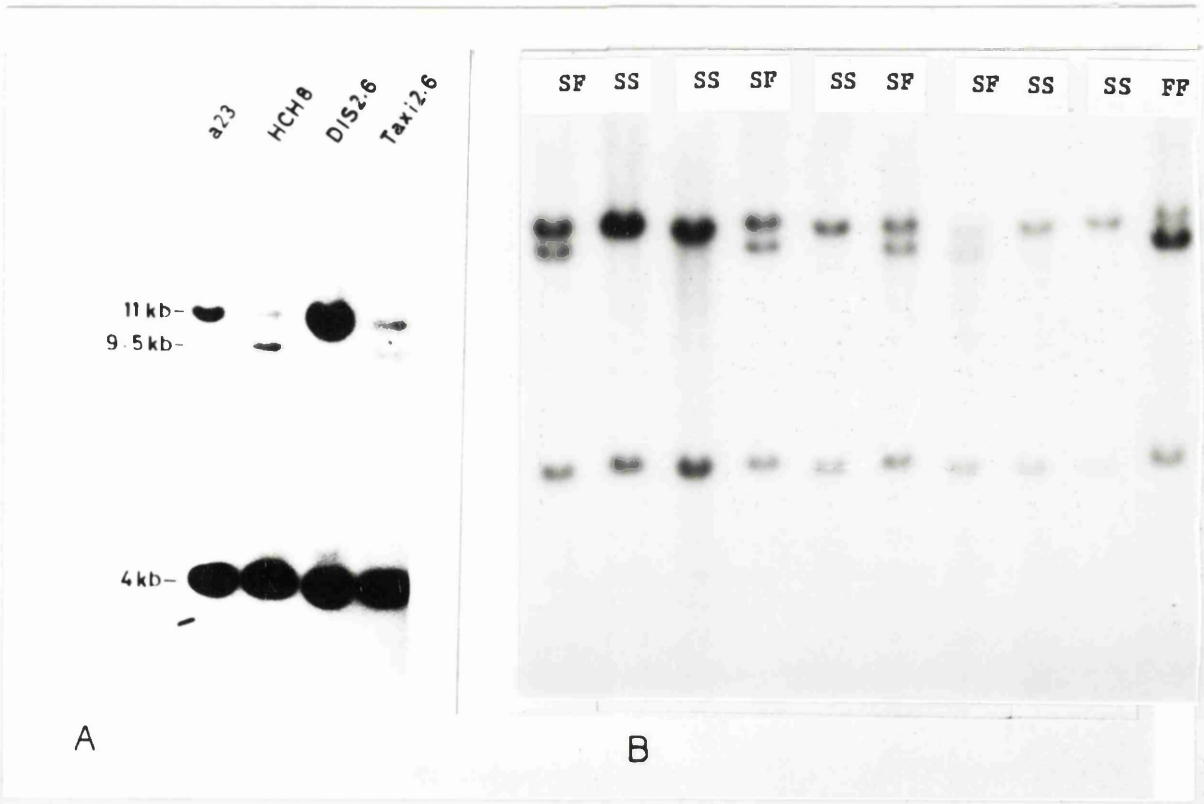


Figure 12. pPGDH1 Detects a BamHI RFLP at the Chromosome 1p Locus.

All DNAs were digested with BamHI and hybridized at high stringency (0.1xSSC/65°C final wash). **Fig.12A** shows the polymorphic 11 and 9.5kb EcoRI bands in three somatic cell hybrids. HCH8 and DIS2.6 were previously shown to contain only the chromosome 1 locus detected H1. HCH8 contains a 9.5kb allele-bearing human chromosome 1 fragment and DIS2.6 possesses a 11kb allele. The 11kb and 4kb constant bands could not be assigned because of comigrating hamster bands.
Fig.12B shows examples of the three RFLP phenotypes where S represents the 11kb allele and F the 9.5kb allele. FF and SF individuals can easily be distinguished despite the 11kb constant band.

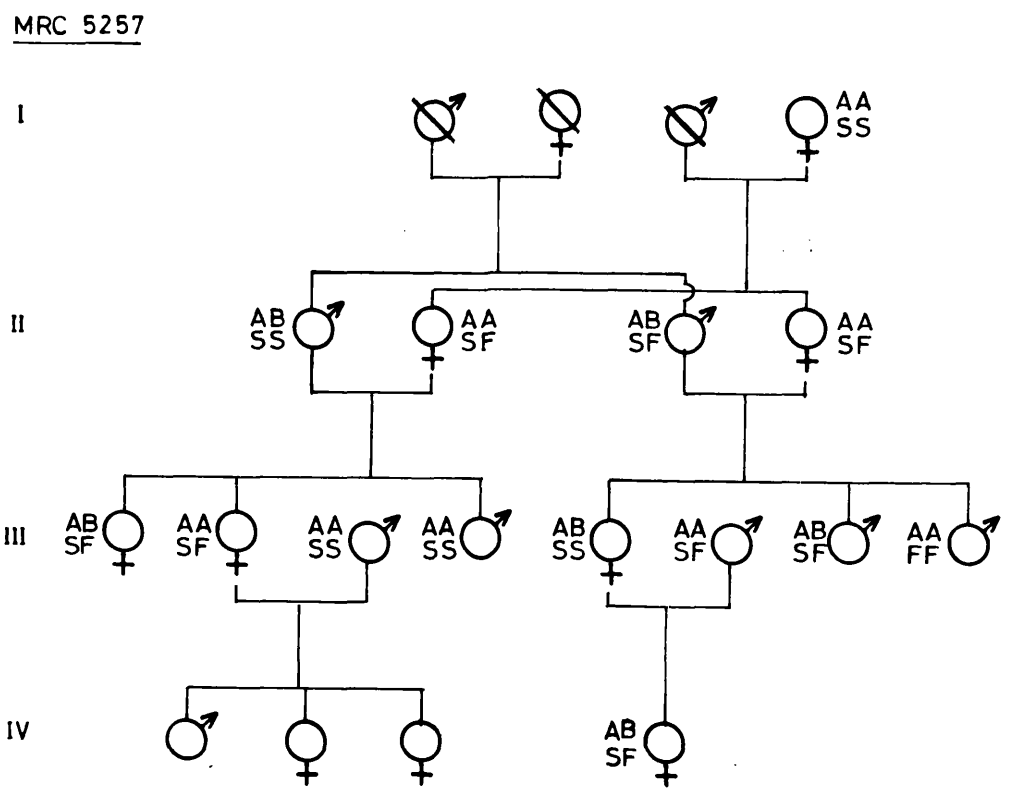


Figure 13. Pedigree Demonstrating the Mendelian Inheritance of the BamHI RFLP Detected by H1

The diagram shows the Mendelian inheritance of the RFLP through four generations of a family. S represents the 11.0kb allele and F the 9.5kb allele. The PGD enzyme electrophoretic phenotypes of the individuals sampled are also shown.

3.8 Analysis of Linkage Between the BamHI RFLP and the PGD Enzyme Polymorphism

This analysis was undertaken to determine the recombination fraction between the two markers as a means of testing the authenticity of H1.

Initially, eight families known from previous studies to segregate for the most common PGD enzyme polymorphism (PGD*B) were sampled (bloods collected by Dr.G.Corney). Samples were typed for the RFLP and the enzyme polymorphism as descibed in the methods. Fig 14 shows an example of a starch gel analysis. Lanes 1 and 2 are AA and AB controls respectively.

Of the eight families, only four were informative (i.e. doubly heterozygous in one parent) and although none of the nine informative meioses showed crossing-over between the two markers, the maximum cumulative lod score (at $\Theta=0$) was still only 1.806 (see Table 3). The lod score at a given recombination fraction is proportional to the number of meioses analyzed and further pegigrees were therefore required to achieve formally significant results. This opportunity was provided by a consideration of the incidence of PGD electrophoretic variants in different populations.

Whereas the frequency of the commonest electrophoretic variant of PGD (PGD*B) is only 0.05 in most European populations, it is significantly higher in a number of non-caucasian ethnic groups (see Appendix A). For example, some native populations of South Africa

have been found, by Prof. Jenkins to have high frequencies of the PGD C allele (Jenkins & Nurse,1974). The parents of eight nuclear Kung families (previously typed by Prof, Jenkins and known to segregate for the enzyme polymorphism) were typed for the RFLP. Only two families were informative (SA1 & SA2) and the results of their analysis are also shown in Table 3. The maximum total lod score for these two families is 1.204 at a recombination fraction of zero (Table 3). Being logarithmic functions of likelihood, lod scores from different data sets may legitimately be added together; combining the Kung data with the lod score for the European pedigrees gives a total lod score of 3.010 at Θ=0 (Table 3).

Linkage is generally considered significant when the lod score exceeds 3.0. Classically, lod scores compare the likelihood of obtaining the observed data at various recombination fractions to a null hypothesis of no linkage (Θ=0.5). In the present case, however, we expect there to be some linkage between the markers since they have both been assigned to 1p36.13-1pter. Therefore, of more interest is the lod score at Θ=0 relative to that at greater values of θ. This is reflected in the graph in Figure 15 which shows a plot of lod scores at different recombination fractions. In fact one could calculate the logarithm of odds of obtaining the pedigree data at any value of Θ relative to any other. For example, a value of Θ=0.06 reflects the

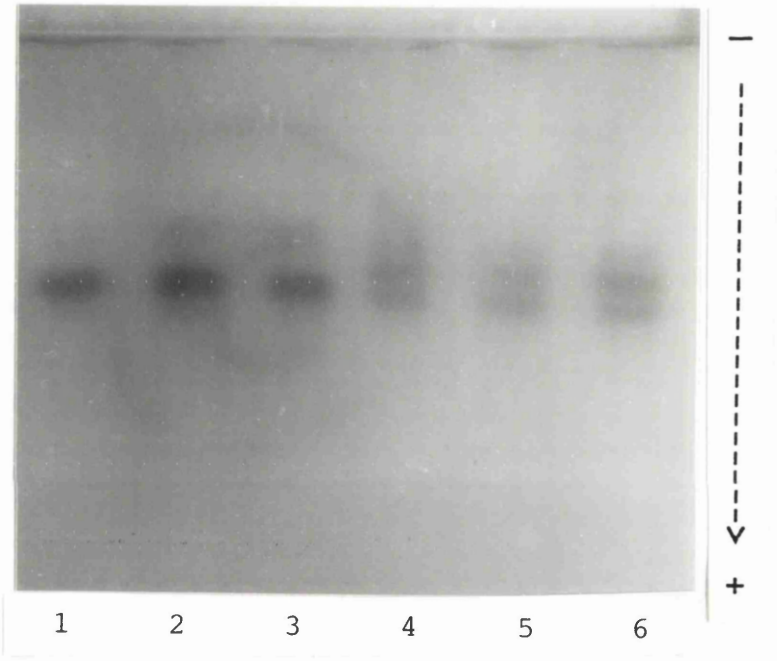


Figure 14. PGD Enzyme Starch Gel.

Saline-washed red cell lysates were run on 11% starch gels in 0.01M, pH7 phosphate buffer. Gels were run for 24 hours at 50V at 4°C prior to slicing and staining (see Methods). Lanes 1-3 are lysates from AA individuals and lanes 4-6 are from AB individuals.

	MALE				FEMALE				
PEDIGREES	NR	R	Z1	LOD(Θ=0)	NR	R	Z1	LOD(Θ=0)	TOTAL LOD(Θ=0)
MRC 160	-	-	2	0.301	-	-	-	-	0.301
MRC 1489	2	0	-	0.602	-	-	-	-	0.602
MRC 4120	-	-	2	0.301	-	-	-	-	0.301
MRC 5257	2	0	-	0.602	-	-	-	-	0.602
SA1	-	-	2	0.301	-	-	2	0.301	0.602
SA2	-	-	-	-	-	-	3	0.602	0.602
TOTAL LOD(Θ=0)				2.107				0.903	**3.010**

TABLE 3. Cumulative Maximum LOD Scores for Linkage between H1 and PGD.

The Lod scores at θ=0 are given for each family. Lod scores obtained from male and female meioses are shown separately, but are summed to give the total maximum Lod score of 3.010. The MRC families are Caucasian whilst SA1 and SA2 are the South African Xhosa pedigrees.
Symbols used: NR - Non-recombinants. R - Recombinants. Z1 - Z1 score.

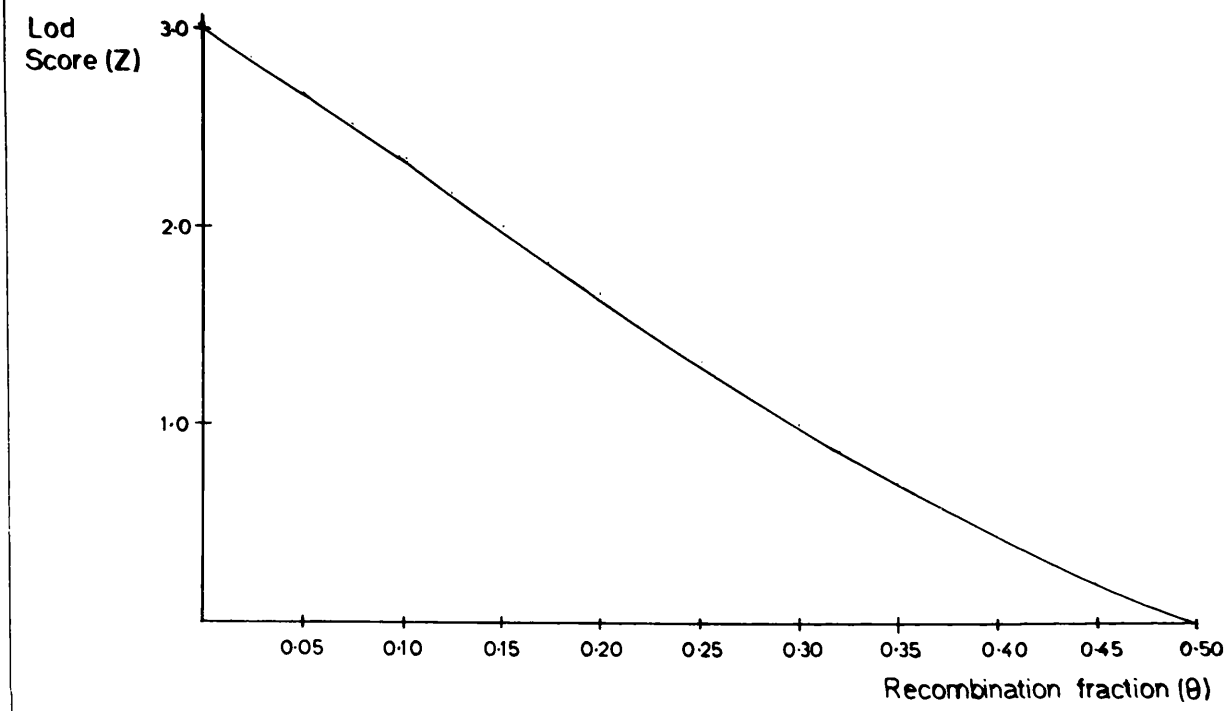


Figure 15. LOD score vs. Recombination Fraction.

The graph shows the total lod score (Z) obtained from the six pedigrees analyzed for all values of the recombination fraction (θ) between the RFLP detected by H1 and the PGD enzyme polymorphism. The maximum Lod score achieved (3.01) is at θ=0 since no crossovers were detected. By defintion Lod (θ=0.5)=0.

approximate maximum genetic distance of chromosome 1p included in the hybrid, DIS 2.6, to which human PGD protein and H1 have both been assigned (see section 3.5.1). In this case the log of odds (Z^*) at $\Theta=0$ relative to $\Theta=0.06$ can be calculated as follows:

$$Z^*(\Theta) = \log_{10}[L(\Theta)/L(0.06)]$$

where L is the likelihood of the observation at the given recombination fraction. For the data collected this works out at $Z^* = 0.18$. The log of odds obtained when comparing different values of Θ clearly depend on the shape of the curve in Fig.15. Thus, this analysis is formally related to the expression of confidence limits in a lod score calculation. These calculations are rather speculative since thay are based on a rough estimate of $\Theta=0.06$ for the comparison. The tight linkage between H1 and the PGD enzyme polymorphism does nevertheless provide support independent of the somatic cell hybrid data for the notion that H1, a cDNA homologous to rat PGD, is in fact human PGD cDNA.

3.9 Analysis of the Allelic Association between the Two Markers.

In order to investigate the question of possible allelic association, the frequencies of all the different genotypes for the two markers found in unrelated individuals were tabulated. In this case 31 unrelated individuals from the European population were typed for both markers. The two AA/FF individuals were included with the AA/SF scores so as to avoid having a row with small numbers subject to large sampling variation (No AB/FF individuals were found). The sample is not random with respect to the enzyme polymorphism since it includes a single AB individual from each of the pedigrees tested during linkage analysis. However, given the null hypothesis of no association between the two markers, this will not bias the sample with repect to RFLP types. The results are given in the table below:

	AA	AB	**TOTALS**
SS	11 (13.5)	8 (5.5)	19
SF+FF	11 (8.5)	1 (3.5)	12
TOTALS	22	9	31

The 'expected' numbers of each cell are in parentheses and are simply calculated by dividing the

product of the marginal totals by the grand total. For example, the 'expected' number of AA/SS individuals is equal to (22x19)/31 = 13.5.

The χ^2 for this table is calculated using the following formula:

$$\chi^2 = \Sigma\ (O-E)^2/E$$

where O is the observed value and E is the expected value. In this case $\chi^2 = 4.12$ with one degree of freedom. This gives a P value of less than 0.05. Thus, the probability of obtaining these results is less than 5% if there is no association between the RFLP and the enzyme polymorphism.

One explanation for allelic association of this sort would be that the markers are in linkage disequilibrium. i.e. that two of the possible four haplotypes are present in the population at higher frequencies than their allele frequencies would suggest. Obviously, knowledge of linkage phase (coupling or repulsion) is required for the construction of haplotypes. This was available for eight unrelated AB individuals, either through the analysis of three-generation families or because of homozygosity at the RFLP locus. All eight B allele-bearing chromosomes were found to possess the S allele also. The chance of this occurring under the assumption of free association of alleles is equal to $(0.76)^8 = 0.1$ where 0.76 is the S allele frequency in the population and thus the chance of a chromosome chosen at random bearing this

allele. This indicates that the markers are in linkage disequilibrium.

In all 9 unrelated individuals from the South African population in which haplotypes could be deduced, the B allele was again present on an S allele-bearing chromosome suggesting a similar departure from equilibrium. However, this effect cannot be quantified since no estimate of the S allele frequency is available for this population.

3.10 Genomic Sequences of Human PGD

H4 was used to screen 4 genome equivalents of a human genomic phage library (Lawn et al.1986). A single clone, λP1, was isolated. This does not hybridize to H3. The genomic DNA used in the construction of this library was partially digested with HaeIII and Sau3A. Thus, the terminal EcoRI sites of the insert were created with linkers and are not present in the genome.

H4 detects two EcoRI fragments in λP1; one of 10.5kb and one of 4.5kb. This correlates with the 10.5 and 4.5kb fragments detected by H4 in Southerns of human genomic DNA cut with EcoRI. These were shown to segregate with the chromosome 1 locus (see section 3.6.1). The 10.5 and 4.5kb EcoRI fragments of λP1 were isolated and subcloned into pAT153 (pPGDE10 and pPGDE4). This allowed a finer restriction map of this region to be constructed (Figure 16).

The 1.4kb PvuII fragment in E10 indicated in Figure 16 was shown to be single copy (by hybridisation of labelled total genomic DNA to blots of PvuII digests of E10) and was used to probe PvuII digests of human, DIS 2.6 hybrid and hamster DNA (Figure 17- experiment carried out by S.Xenophontos, Galton Laboratory). A 1.4kb fragment in both DIS2.6 and human DNA demonstrates not only the existence of P1.4 in genomic DNA but also that λP1 represents the PGD gene itself and not a pseudogene since DIS 2.6 had previously been shown to contain only the chromosome 1 H4-hybridizing

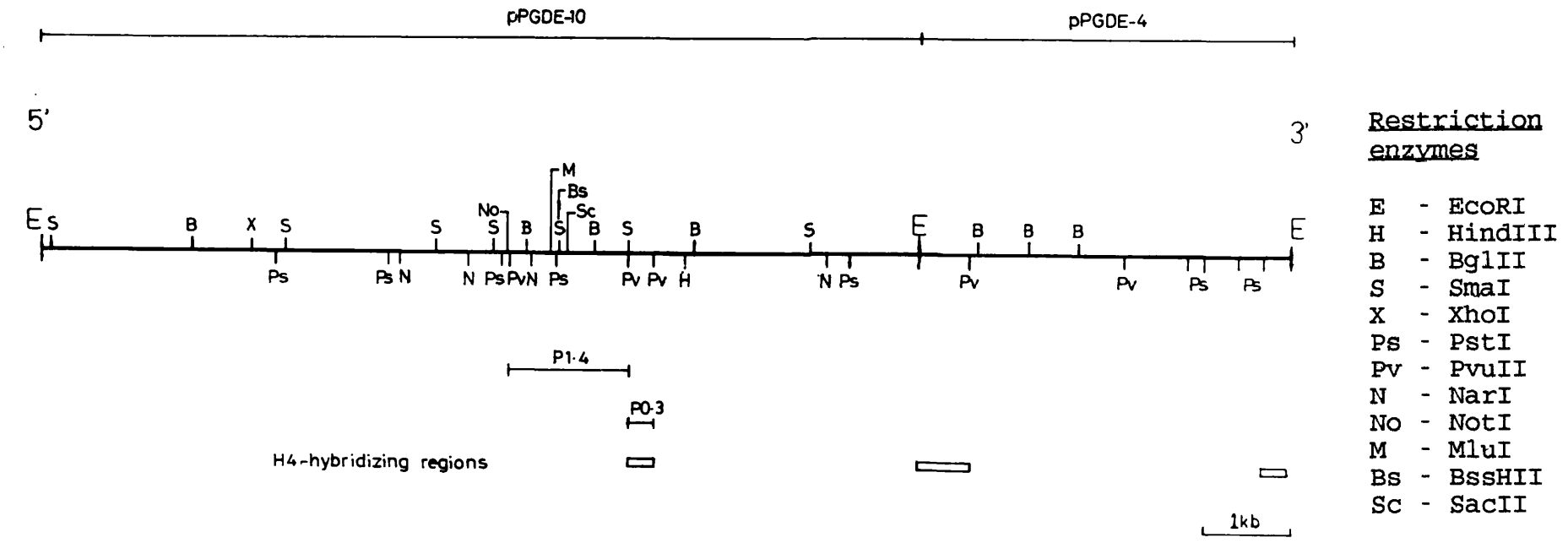


Figure 16. Restriction maps of pPGDE10 and pPGDE4.

The 10.5 and 4.4kb EcoRI fragments of λP1 (which hybridize to H4) were subcloned and mapped with the restriction enzymes listed above. Hybridisation of H4 to partial EcoRI digests of λP1 showed that E10 and E4 are adjacent to each other in the phage. P1.4 is the single-copy fragment containing part of the CpG island. All of the E10 sequences homologous to H4 lie within the PvuII fragment P0.3. All sites for the enzymes listed are shown except for SmaI which was not used to map pPGDE4.

Figure 17. Hybridisation of P1.4 to Human, Hydrid and Hamster genomic DNA

All DNAs were digested with PvuII.

P1.4 is the single copy PvuII fragment isolated from the genomic clone λP1 (see fig.16). Lane 1 (human) shows that the 1.4kb PvuII fragment does indeed exist in the genome.

DIS2.6 is a hybrid which contains the human PGD locus. It does not contain any other H4-related locus. Thus, the detection of a 1.4kb PvuII fragment by P1.4 in DIS2.6 (and its absence in the parent hamster line,a23) confirms that λP1 originates from the PGD gene rather than from related sequences detected by H4.

The considerable non-specific hybridisation is probably due to contamination of the probe with repetitive sequence DNA.

locus. This was confirmed by hybridizing H4 to digests of pPGDE10 and pPGDE4. Three different H4-hybridizing regions were delineated as shown in Figure 16 indicating that adjacent exons in the cDNA are separated in the genomic clone by introns. The DNA separating these H4-hybridizing regions can be said to be exclusively intronic. Comparison of the restriction maps of the genomic clones with the restriction map of H4 (obtained by restriction digests and sequence analysis), regions of homology between the cDNA and genomic clones were located allowing the determination of the orientation of the gene in λP1 and providing preliminary information concerning the intron/exon structure of the gene. For example, BglI and BalI sites, 61bp and 55bp from the 5' end of H4 respectively (see later, fig.22), mapped very close to each other within the 0.3kb PvuII fragment of E10 (Fig.16; BglI and BalI sites not shown because the other sites where these cleaved were not mapped). This established that the orientation of the gene is as indicated in figure 16.

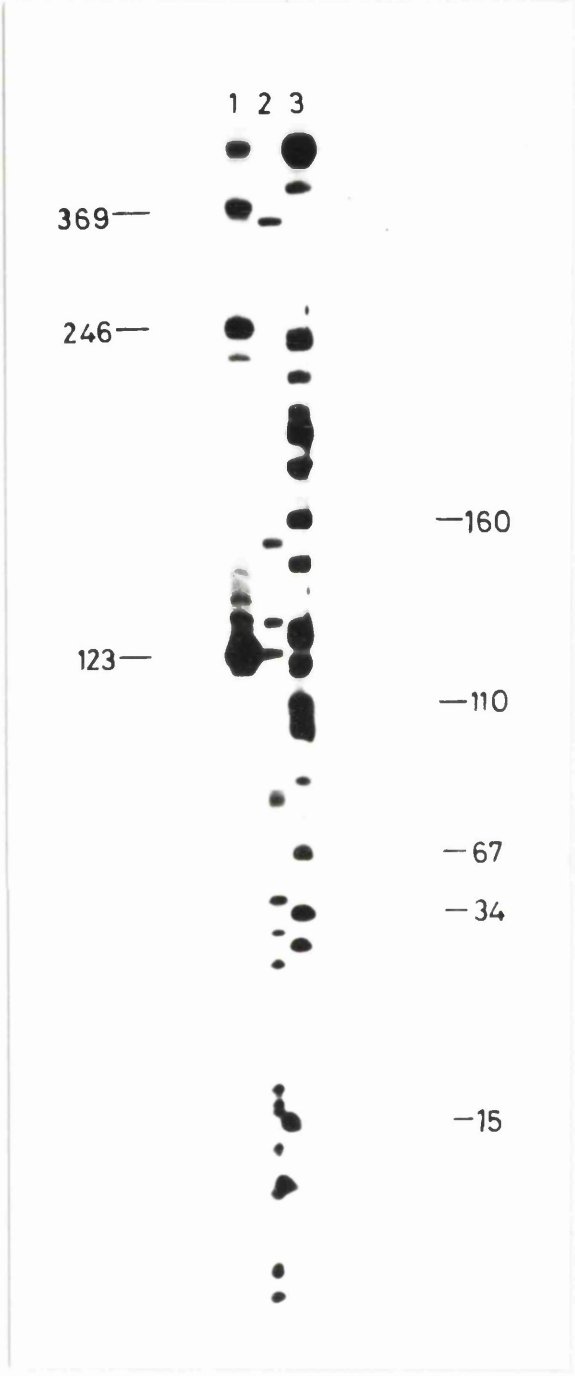
3.11 A CpG Island Associated with the PGD Gene

The restriction map of E10 shows a cluster of sites for BssHII, SacII, MluI, NotI and NarI (see figure 16). All of these enzymes recognize target sequences which either contain only G and C nucleotides or are GC-rich. These sequences are rare partly because of the low G+C content in mammalian genomes. Furthermore, in each case, within these sequences is at least one copy of the dinulcleotide CpG which is under-represented in mammalian genomes. This is probably a consequence of cytosine methylation. Cytosine is prone to deamination, and when methylated this gives rise to thymine. This results in a deficiency of CpG and an over-representation of TpG and CpA in the genome. Some small regions (0.5-3kb) of the genome, however, have been found to have a high G+C content (about 65%) and a correspondingly high frequency of CpG dinucleotides. These sequences are found to be hypomethylated in the genome and the absence of CpG under-representation is thought to be a consequence of this. The regions are known as CpG or HpaII tiny fragment (HTF) islands. The great interest in CpG islands stems from their association with genes (Bird,1986). The presence of a cluster of sites for these restriction enzymes with rare target sequences (`rare-cutters') is thus indicative of the presence a CpG island. Indeed, the strong association between CpG islands and genes enables genes to be detected using these rarely-cutting

restriction enzymes.

In order to discover in more detail the distribution of CpG's in this region, the PvuII fragment encompassing most of these sites was analyzed with HpaII. The PvuII fragment (P1.4, see fig.16) was gel-purified, digested with HpaII, end-labelled with Klenow fragment, denatured and run on a sequencing gel (see Methods). The results of this analysis are shown in figure 18. The markers used were a 123bp ladder (lane 1) and pAT153 cut with HpaII (lane 3). Figure 18 shows that P1.4 is cut into at least 14 fragments and all but one are 160bp or less.

The only large fragment is one of 360bp. This can be partially mapped by inspection of the distribution of SmaI sites (see figure 16). The target sequence for SmaI (CCCGGG) contains within it the target sequence for HpaII and therefore HpaII will cut at every SmaI site. The location of one of the two SmaI sites within P1.4 indicates that the 360bp fgagment cannot lie at the 5' end of P1.4 as drawn in figure 16. Also, the distribution of the other rare-cutter sites suggests that the middle region of P1.4 is GC-rich and suggests that the 360 bp fragment lies at the 3' end of P1.4. If this is true, the remaining 1040bp section of P1.4 is cut into 13 fragments indicating an average size of 80bp. Such a spacing for the HpaII target site (CCGG) would be expected for a DNA sequence with a G+C content of 67%. This is very close to the average G+C content



Lanes:

1.123bp ladder
2.P1.4 x HpaII
3.pAT153 x HpaII

Figure 18. Analysis of CpG Distribution within the 1.4kb PvuII fragment of pPGDE10.

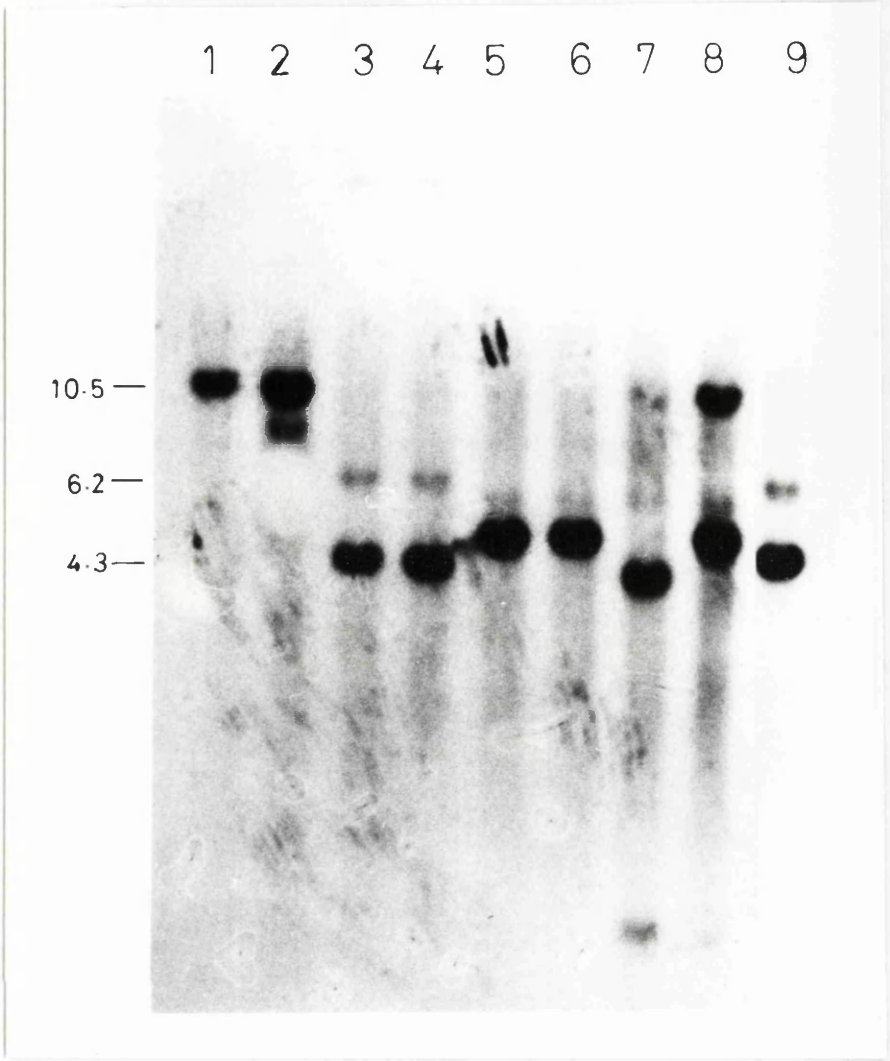
P1.4 was gel-purified and digested to completion with HpaII. The products were end-labelled and denatured prior to separation on a 6% urea polyacrylamide gel (see Methods). P1.4 is cleaved into at least 14 fragments by HpaII. All but are less than 160 bp. Marker sizes are indicated in nucleotide bases.

of 65% found in the CpG islands analyzed to date (Bird,1986). Thus the CpG island probably begins 360bp into P1.4 and extends for at least 1kb.

The location of the CpG island with respect to the H4 hybridizing-regions of E10 and E4 is consistent with the orientation of the gene such that the CpG island marks the 5' end of the gene and possibly extends into the first exon as has generally been found to be the case (Bird,1987).

As well as a high G+C content and a proportionate representation of CpG dinucleotides, the other criterion for the identification of CpG islands is their hypomethylation in the genome. Whereas most of the mammalian genome is methylated, CpG islands are characterized by lack of methylation (Bird,1986). Whether this is in some way a consequence of the frequent binding of transcription factors or that, more interestingly, this serves as a marker for the presence of a gene to such factors has not yet been established. Investigations into the influence of methylation in promoter regions have so far yielded contrasting results (Dynan ,1989).

The state of methylation of G+C-rich region associated with PGD was investigated by digesting genomic DNA with EcoRI together with a variety of rare-cutters which were shown to cut within E10 (see fig.16). All of these enzymes are sensitive to cytosine methylation. These digests were probed with P1.4 as shown in fig.19



Lanes:

1. **EcoRI**
2. **ERI + XhoI**
3. **ERI + BssHII**
4. **ERI + SacII**
5. **ERI + NotI**
6. **ERI + NotI**
7. **ERI + SmaI**
8. **ERI + NarI**
9. **ERI + MluI**

Figure 19.Hybridisation of P1.4 to Human Genomic DNA Digested with EcoRI and Rare-Cutters

This experiment was performed to investigate the state of methylation at the putative CpG island within E10. All of the restriction enzymes used (except EcoRI) are methylation -sensitive. All of these recognize sites within the CpG island in cloned DNA and are shown to cleave at the same positions in genomic DNA demonstrating that these sites are not methylated *in vivo*. The DNA used was from white blood cells.

Lane 2 shows that the XhoI site is at least partially methylated since the 10.5kb EcoRI site is still clearly visible. The same is the case for NarI (lane 8). It may also be that one or both of these enzymes has not been fully active for trivial technical reasons. It appears that two SmaI sites in the CpG island are unmethylated since a small band is visible in lane 7.

continued on opposite page

(this experiment was carried out by S.Xenophontos, Galton Laboratory). Lane 1 shows the the 10.5 kb fragment detected by P1.4 for comparison. Lanes 3-9 show that NotI, BssHII, MluI, SacII, SmaI and NarI cleave the 10.5kb fragment demonstrating that the G+C-rich region of E10 forfills all the criteria of a CpG island. The site for XhoI lies outside the CpG island and is certainly partially protected from cleavage (lane 2). SmaI (lane 7) cuts twice within the island since a small band can be seen in this digest.

The detection of a 6kb EcoRI fragment at the 1p locus (see figure 9) by H4 clearly demonstrates that the PGD gene extends beyond λP1. This fragment probably lies adjacent to E4 in the genome although the theoretical possibility that one or more exclusively intronic EcoRI fragments (of any size) lie between these two fragments cannot be excluded since these would not hybridize to cDNAs. H3 and H4 are contiguous segments of cDNA and therefore the 3.5kb EcoRI fragment detected by H3 must be immediately 3' of the 6.0kb EcoRI fragment detected by H4. Thus, the tentatative map of the PGD gene can be constructed as illustrated in figure 20. This shows that the gene covers about 21kb of the genome from the end of 3' untranslated region to the region containing the CpG island.

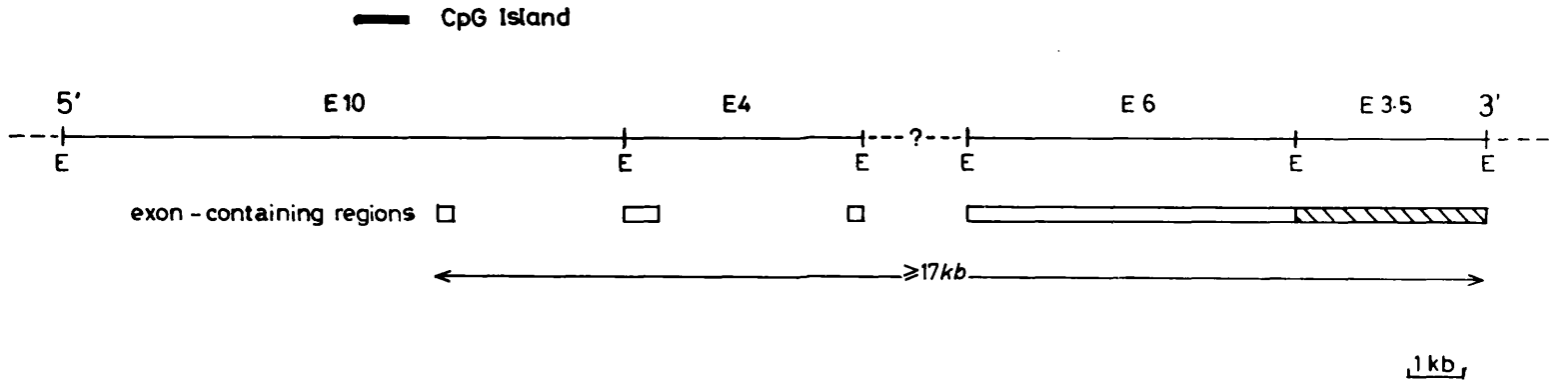


Figure 20. Putative Structure of the Human PGD Locus

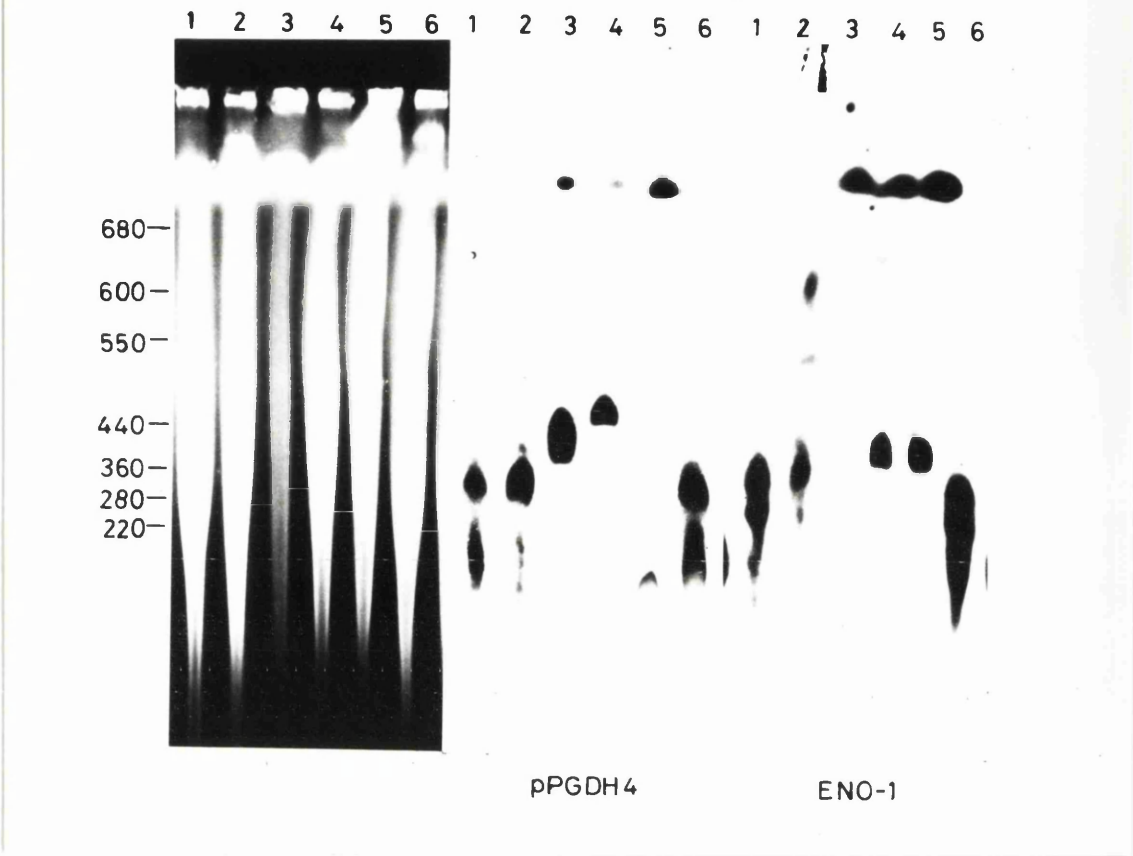
The figure shows an EcoRI restriction map of the PGD locus deduced from the hybridisation patterns of the cDNAs to genomic DNA. The exon-containing regions are those which hybridize to pPGDH3 (hatched box) and pPGDH4 (open boxes).
E10 and E4 have been shown to be adjacent by restriction mapping of λP1 (see fig.16). E6 and E3.5 must be adjacent since they hybridize to H4 and H3 respectively which are adjacent (separated by an EcoRI site) in the cDNA clone, λTC12 (see fig.1). The possibility that E6 and E4 are separated in the genome by intronic sequences cannot be excluded.
The approximate location and extent of the CpG island are shown. The transcribed sequences of the gene span more than 17kb.

3.12 Long-Range Mapping of Region Containing the PGD Locus

An advantage of using a gene probe such as H4 for long-range physical mapping is its association with CpG islands. The restriction enzymes which cleave rarely and that tend to be clustered in CpG islands are those used for mapping by pulse-field gel electrophoresis (PFGE). Therefore the position of a gene probe within a hybridizing PFGE fragment can be determined. For example, from the restriction map of E10, it can be seen that H4 detects sequences close to one end of SacII, MluI, BssHII, NotI and NarI fragments. These enzymes are therefore useful in establishing a long range map downstream of the PGD locus when their digests are probed with H4. A single-copy probe from E10 upstream of the CpG island would of course provide mapping data on the other side of the PGD locus.

Figure 21 shows genomic DNA plus digested with a variety of rarely-cutting restriction enzymes and run under conditions which resolve DNA fragments of about 100-750kb (170V/50sec pulse-time/24 hour run/0.8% agarose/5° C).

H4 detects a 440kb MluI fragment (lane 3) which is cleaved by NarI (lane 2) to 350kb. A 350kb fragment is also detected in DNA cut only with NarI (lanes 1 & 6) indicating that H4 detects a 350kb NarI fragment with no sites for MluI. Unfortunately, since H4 detects at



Lanes:
1. NarI
2. NarI+MluI
3. MluI
4. SalI
5. SacII
6. NarI

Figure 21. Long-Range Physical Mapping of PGD and ENO1 Loci

DNA blocks were digested with the restriction enzymes listed above and run on a 0.8% agarose pulse-field gel (170V/pulse-time=50 seconds/24 hours/5°C). The blotted DNA was first hybridized to pPGDH4. After autoradiography, the probe was washed off and the filter rexposed to ensure complete removal of the probe prior to rehybridisation to ENO1. Both autoradiographs shown were exposed to the filter for one week. The sizes indicated are in kb and correspond to yeast chromosomes used as size markers.

Both probes hybridize to 350kb NarI fragments (lanes 1 and 6) which are not cleaved by MluI (lane 2). The hybridisation of H4 to a larger MluI fragment (lane 3) is consistent with this. However, no ENO1-hybridizing MluI fragment was resolved and the two probes hybridize to different SalI and SacII fragments.

least one locus other than the PGD gene, these bands cannot be assumed to originate from the 1p locus.

In the case of the 350kb NarI fragment indirect evidence for its localisation to 1p has been obtained. The same blot (after having the H4 probe removed and been tested by autoradiography) was probed with a full-length cDNA probe for Enolase-1 (pSA082; Giallongso et al. 1986). Eno-1 has been assigned to the same region of the genome as PGD (1p36.13-1pter). The most recent estimate of the genetic distance between Enolase-1 and PGD is 1.5cM (Morton,1988; Keats et al.1989). The result of the hybridisation of pSA082 to the same PFG blot to which H4 was hybridized is shown in fig.21. Between successive hybridisations, the PFG blot was stripped of probe (see Methods) and the filter was autoradiographed to ensure no signal remained on the filter. The enolase probe also detects a 350kb NarI fragment (lanes 1 &6) which again is left uncut after digestion with MluI (lane 2). This suggests that the enolase-1 and PGD loci may exist on a common 350kb NarI fragment. Since the MluI fragment detected by H4 is larger than the NarI fragment and an MluI site exists in the PGD CpG island, this fragment should also encompass both loci if it originates from chromosome 1p. Unfortunately no MluI fragment which hybridizes to enolase is resolved. The two probes hybridized to differently-sized SalI and SacII fragments. Since H4 detects at least two unlinked loci, no firm conclusions

can be drawn on the basis of these results.

Thus, at present the physical linkage between PGD and ENO1 on a 350kb fragment is still tentative. If it proves to be correct, it establishes the orientation of the PGD gene on the chromosome. NarI cleaves within the CpG island as demonstrated by the experiment shown in figure 19, and thus the 350kb fragment must be bounded by this site. Since H4 hybridizes to the 3' side of this site in the PGD gene, the other end of this fragment and ENO1 must lie beyond the 3' end of the PGD gene. The ENO1 locus has been shown to be distal of PGD on chromosome 1p (Carritt et al.1982) and therefore the direction of transcription of the PGD gene must be towards the telomere.

3.13 Sequence Analysis of cDNAs.

All four cDNAs were subcloned into M13mp18 in order to obtain sequence information for the following purposes:

1. To establish the existence and the extent of an open reading frame.
2. To determine the direction of transcription.
3. To establish the extent of the homology between H1, H2 and H3.
4. To determine the orientation of the cDNA relative to the genomic clone λP1.
5. To test for sequence homology between human PGD and the *E.coli* PGD gene which has been cloned and sequenced (Nasoff et al.1984/K02072) as well as the sheep sequence PGD polypeptide sequence which has also been determined (Carne & Walker,1983).

Sequencing of H4 and H3 did indeed reveal an open reading frame (ORF) which includes the whole of H4 and ends within H3 (see figure 22).

The orientation of the ORF and alignment of the restriction sites detected by sequencing with the previously constructed map λTC12 (figure 1) confirmed that H4 is 5' of H3. The sequences of H1 and H2, in keeping with the restriction mapping results, show that H1, H2 and H3 share a common 3' end and that H2 and H3 are identical at both ends.

5'- H4

```
                                                                    42
GTC TGT GCT TTT AAT AGG ACT GTC TCC AAA GTT GAT GAT TTC
Val Cys Ala Phe Asn Arg Thr Val Ser Lys Val Asp Asp Phe

     Ba       Bg                                                    84
TTG GCC AAT GAG GCA AAG GGA ACC AAA GTG GTG GGT GCC CAG
Leu Ala Asn Glu Ala Lys Gly Thr Lys Val Val Gly Ala Gln

                                                                   126
TCC CTG AAA GAG ATG GTC TCC AAG CTG AAG AAG CCC CGG CGG
Ser Leu Lys Glu Met Val Ser Lys Leu Lys Lys Pro Arg Arg

                                                                   168
ATC ATC TCC TGG GTG AAG GCT GGC AAG CTG TGG ATG ATT TCA
Ile Ile Ser Trp Val Lys Ala Gly Lys Leu Trp Met Ile Ser

                     K                                             210
TCG AGA AAT TGG TAC CAT TGT TGG ATA CCT GGT GAC ATC ATC
Ser Arg Asn Trp Tyr His Cys Trp Ile Pro Gly Asp Ile Ile

                                                                   252
ATT GAC GGA GGA AAT TCT GAA TAT AGG GAC ACC ACA AGA CGG
Ile Asp Gly Gly Asn Ser Glu Tyr Arg Asp Thr Thr Arg Arg

                                                        287
TGC CGA GAC CTC AAG GCC AAG GGA ATT TTA TTT GT---------
Cys Arg Asp Leu Lys Ala Lys Gly Ile Leu Phe

------------about 200 bases not sequenced--------------

                                                                    42
ATA GAG TAT GGC GAC ATG CAG CTG ATC TGT GAG GCA TAC CAC
Ile Glu Tyr Gly Asp Met Gln Leu Ile Cys Glu Ala Tyr His

                                                                    84
CTG ATG AAA GAC GTG CTG GCA TGG CGC AGA CGA GAT GCC CAG
Leu Met Lys Asp Val Leu Ala Trp Arg Arg Arg Asp Ala Gln

S                                                                  126
GCC TTT GAG GAT TGG AAT AAG ACA GAG CTA GAC TCA TTC CTG
Ala Phe Glu Asp Trp Asn Lys Thr Glu Leu Asp Ser Phe Leu

                                                                   168
ATT GAA ATC ACA GCC AAT ATT CTC AAG TTC CAA GAC ACC GAT
Ile Glu Ile Thr Ala Asn Ile Leu Lys Phe Gln Asp Thr Asp

                                                                   210
GGC AAA CAC CTG CTG CCA AAG ATC AGG CGA GCT GCG GGG GCT
Gly Lys His Leu Leu Pro Lys Ile Arg Arg Ala Ala Gly Ala

                                                                   252
AAG AAG GGC ACA GGG AAG TGG ACC GCC ATC TCC GCC CTG GAA
Lys Lys Gly Thr Gly Lys Trp Thr Ala Ile Ser Ala Leu Glu
```

Continued overleaf

Figure 22 DNA and Amino Acid Sequence of λTC12

See overleaf for legend

```
                                                                  294
TAC GGC GTA CCC GTC ACC CTC ATT GGA GAA GCT GTC TTT GCT
Tyr Gly Val Pro Val Thr Leu Ile Gly Glu Ala Val Phe Ala

                                  <---H4   E   H3--->             336
CGG TGC TTA TCA TCT CTG AAG GAT GAG AGA ATT CAA GCT AGC
Arg Cys Leu Ser Ser Leu Lys Asp Glu Arg Ile Gln Ala Ser

                                                                  378
AAA AAG CTG AAG GGT CCC CAG AAG TTC CAG TTT GAT GGT GAT
Lys Lys Leu Lys Gly Pro Gln Lys Phe Gln Phe Asp Gly Asp

         X                                                        420
AAG AAA TCA TTC CTG GAG GAC ATT CGG AAG GCA CTC TAC GCT
Lys Lys Ser Phe Leu Glu Asp Ile Arg Lys Ala Leu Tyr Ala

                                                                  462
TCC AAG ATC ATC TCT TAC GCT CAA GGC TTT ATG CTG CTA AGG
Ser Lys Ile Ile Ser Tyr Ala Gln Gly Phe Met Leu Leu Arg

                                                                  504
CAG GCA GCC ACC GAG TTT GGC TGG ACT CTC AAT TAT GGT GGC
Gln Ala Ala Thr Glu Phe Gly Trp Thr Leu Asn Tyr Gly Gly

                                                                  546
ATC GCC TTG ATG TGG AGA GGG GGC TGC ATC ATT AGA AGT GTA
Ile Ala Leu Met Trp Arg Gly Gly Cys Ile Ile Arg Ser Val

                                                                  578
TTC CTA GGA AAG ATA AAG GAT GCA TTT GAT CGA AAC CCG GAA
Phe Leu Gly Lys Ile Lys Asp Ala Phe Asp Arg Asn Pro Glu

                                                                  630
CTT CAG AAC CTC CTA CTG GAC GAC TTC TTT AAG TCA GCT GTT
Leu Gln Asn Leu Leu Leu Asp Asp Phe Phe Lys Ser Ala Val

  655
GAA A--------------about 330 bases not sequenced-------
Glu
                                                                   42
CCC TAT TTT CTG TTC AGT TTT TTA AAA GTG TTG TAA GAG ACT CCT GAG
                                                                   84
GAA GAC ACA CAG TTT ATT TGT AAA GTA GCT CTG TGA GAG CCA CCA TGC
                                                                  126
CCT CTG CCC TTG CCT CTT GGG ACT GAC CAG GAG CTG CTC ATG TGC GTG
                                                                  168
AGA GTG GGA ACC ATC TCT TGC GGC AGG TTG GCT TCC GCG TGC CCC GTG
                                                                  210
TGC TGG TGC GGT TCC CAT CAC GCA GAC AGG AAG GGT GTT TGC GCA CTC
                                          E     247
TGA TCA ACT GGA ACC TCT GTA TCA TGC GGC TGA ATT C- H3-3'
```

Figure 22. DNA and Amino Acid Sequence of λTC12

The inferred amino acid sequence and the coding DNA strand of the sequenced regions of the cDNA clone λTC12 are shown. Restriction sites are indicated by underlined sequence (E-EcoRI;Ba-BalI;Bg-BglI;P-PvuII;S-StuI;X-XmnI;K-KpnI). The BalI site overlaps one of the BglI sites. The stop codons in all three reading frames in the 3'sequenced region of H3 are shown in bold type. The open reading frame includes all of H4 and ends in H3.

The ORF continues in the 5' region of H3, but it ends within this subclone since the 3' sequence of H3 contains numerous stop codons in all three reading frames (fig.22). This is consistent with the earlier prediction that H1 is, in large part, made up of 3' untranslated sequences.

The DNA sequence obtained from H3 and H4 was compared to the published full-length sequence of the *E.coli* PGD gene (Nasoff et al.1984; GENBANK database no.K02072). This was carried out using the "SEQ" and "PEP" program from Intelligenetics Inc.- see Methods). No homology between the 3' region of H3 and the *E.coli* gene was detected as might be expected for untranslated sequences. However, homologous bacterial sequences were detected with the other two sequenced regions of the human cDNA. Homology was detected between the 293 bases sequenced at the 5' end of H4 and 281 bases from the 5' end of the *E.coli* gene (84 bases from the translation start codon). This showed a maximum homology of 48% with a 3bp insertion in the human sequence relative to that of *E.coli* The homology undoubtedly continues beyond the sequenced 5' region of H4 since about 200 bases downstream in H4 (and 201 bases downstream in the *E.coli* gene), homology was again detected. The 670 bases at the the 3' end of H4 and the 5' end of H3 were also homologous to bacterial sequences. In this case the maximum homology was 57% and was achieved by allowing two 3bp insertions in the human sequence and

one 3bp insertion in the *E.coli* sequence. The average homology between the 900 bases of human cDNA and the bacterial gene was 53%. This is obviously significant since two sequences chosen at random will, on average, only show 25% homology.

In order to assess the functional significance of the sequence conservation, the peptide sequences derived from the DNA sequences were compared.

As expected, the conservation between *E.coli* and human PGD sequences at the amino level was even stronger (figure 23). The 96 amino acids encoded at the 5' end of H4 share 40% homology with 95 amino acids at the amino-end of the *E.coli* protein (28 amino acids from the *E.coli* cap site). The second region of homology identified by nucleotide comparison showed 54% homology at the amino acid level. The expected homology between two randomly-chosen peptide sequences is about 5% (assuming them to be made up exclusively of the 20 common amino acids) and therefore the average amino acid homology of 50% between the human and *E.coli* sequences is considerably higher than the 53% homology at the DNA level. This is not surprising since many of the differences at the DNA level involve the third bases of codons which often do not alter the amino acid encoded and therefore will be phenotypically silent and not under the same constraint as nucleotides in the first or second positions of codons.

Figure 23. Comparison of Human, Sheep and *E.coli* PGD Polypeptide Sequences.

The cDNA-derived *E.coli* and human polypeptide sequences are shown aligned with the re-ordered sheep sequence. The numbering refers to the *E.coli* sequence. The stars (*) indicate homology between two sequences. The standard single-letter amino acid code is used:

A-alanine R-arginine N-asparagine D-aspartic acid C-cysteine Q-glutamine E-glutamic acid
G-glycine H-histidine I-isoleucine L-leucine K-lysine M-methionine F-phenylamine
P-proline S-serine T-threonine W-tryptophan Y-tyrosine V-valine

```
                10         20         30         40         49         59         69         79         89         99
E.COLI    MSKQQIGVVG MAVMGRNLAL NIESRGYTVS IFNRSREKTE EVIA-ENPGK KLVPYYTVKE FVESLETPRR ILLMVKAGAG TDAAIDSLKP YLDKGDIIID
                                 *   ***   *      * *  *  * *      **  *  *  *** *   ****                   ******
HUMAN                            VC AFNRTVSKVD DFLANEAKGT KVVGAQSLKE MVSKLKKPRR IISWVKAGKL WMISSRNWYH CWIPGDIIID
                                 ** ********** ********** ***** ** * ********** **  ****                   * ****
SHEEP          AQADL IGLMGEDLIL NMNDHGFVVC AFNRTVSKVD DFLANEAKGT KVLGAHSLEE MVSKLKKPRR IILLVKAGEA VDNFIEKLVP LLDIGSIIID

                109        119        129        139        149        159        169        179        189        199
E.COLI    GGNTFFQDTI RRNRELSAEG FNFIGTGVSG GEEGALKGPS IMPGGQKEAY ELVAPILTKI AAVAEDGEPC VTYIGADGAG HYVKMVHNGI EYGDMQLIAE
          ***    **  ** * * * *   *                                                                * ******** *
HUMAN     GGNSEYRDTT RRCRDLKAKG ILF                                                                I EYGDMQLICE
          *** *****  ******* ** ***                                                                * **********
SHEEP     GGNNEYRDTM RRCRDLKDKG ILFVGSGVSG GEDGARYGPS LMPGGNKEAW PHIKAIFQGI AAKVGTGEPC CDWVGDDGAG HFVKMVHNGI EYGDMQLICE

                207        217        227        237        247        257        266        276        286        296
E.COLI    AYSL--LKGG LNLTNEELAQ TFTEWNNGEL SSYLIDITKD IFTKKDEDGN YLVDVILDEA A-NKGTGKWT SQSALDLGEP LSLITESVFA RYISSLKDQR
          ** *   *   *       **  *   ** ** *  ** **   *    * **   *   *   *    *******   ***  * *   ** * *** *  ***** *
HUMAN     AYHL--MKDV LAWRRR-DAQ AFEDWNKTEL DSFLIEITAN ILKFQDTDGK HLLPKIRRAA GAKKGTGKWT AISALEYGVP VTLIGEAVFA RCLSSLKDER
          ****  **** *       *  *** ****** *********  ****** **  *******  * *  ***** * ********** ********** **********
SHEEP     AYHLDGMKDV LGLGHKEMAK AFEEWNKTEL DSFLIEITAS ILKFQDADGW HLLPKIRDSA G-QKGTGKGT AISALEYGVP VTLIGEAVFA RCLSSLKDER

                306        315        325        335        345        355        365        375        385        395
E.COLI    VAASKVLSGR QA-QPAGDKA EFIEKVRRAL YLGKIVSYAQ GFSQLRAASE EYNWDLNYGE IAKIFRAGCI IRAQFLQKIT DAYAENPQIA NLLLAPYFKQ
            *** * *  *  *  ***   * *  * ** *  ** **** **  ** *   *  * **** **   * *** ** ** **  **   **    ****  **
HUMAN     IQASKKLKGP QKFQFDGDKK SFLEDIRKAL YASKIISYAQ GFMLLRQAAT EFGWTLNYGG IALMWRGGCI IRSVFLGKIK DAFDRNPELQ NLLLDDFFKS
          ********** *   * **** **** ***** ********** * ******** ********** ********** ********** ******* ** **********
SHEEP     IQASKKLKGP QDIPFQGDKK SFLENIRKAL YASKIISYAQ GWMLLRQAAT EFGWTLNYGG IALMWRGGCI IRSVFLGKIK DAFDRNPGLQ NLLLDDFFKS

                405        415        425        435        445        455        464    468
E.COLI    IADDYQQALR DVVAYAVQNG IPVPTFSAAV AYYDSYRAAV LPANLIQAQR DYFGAHTYKR IDKEGVF-HT EWLD

HUMAN     AVE
          ***
SHEEP     AVENCQDSWR RAISTGVQAG IPMPCFTTAL SFYDGYRM-- LPANLIQAQR DYFGAHTYEL LAKPGQFIHT DGSW
```

Figure 23. Comparison of Human, Sheep and *E.Coli* PGD Polypeptide Sequences.

Comparison of the human cDNA-derived amino acid sequence with the sheep sequence obtained from direct sequencing of the protein (Carne & Walker,1983) immediately revealed discrepancies between the two sequences. Closer inspection showed that the sheep sequence contained regions homologous to all regions of the human sequence. However, the regions of homology were in different orders in the two sequences. The strategy of direct sequencing of proteins is particularly prone to ordering errors of this type: the polypeptide is cleaved with a variety of reagents. The smaller peptides produced can then be sequenced by Edman degradation or other methods. Thus, the sequencing protocol of Carne and Walker (1983) was closely examined and it was found that, in each case, the boundaries between homologous segments in the sheep sequence correspond to sites of cleavage of cyanogen bromide, one of the reagents used by Carne and Walker to fragment the polypeptide.

Therefore, I have corrected the ovine sequence by reordering it the following way:

NH_2- 1-9 ; 105-432 ; 84-104 ; 10-83 ; 433-466 -COOH

(the numbering refers to that in Carne & Walker,1983)

All of these segments correspond to partial or complete cyanogen fragments. In fact, the authors acknowledged that no overlapping fragments across the junction between amino acids 104 and 105 were obtained. As for the three other junctions, the sequenced over-

laps were only 1 base (9-10), 4 bases (83-84) and 1 base (432-433).

Together with the finding that the human and *E.coli* amino acid sequences are completely co-linear in their regions of homology, this suggests that, with little doubt, it is the ovine sequence which is erroneous. For the purposes of the comparison, the corrected sheep sequence was used. This sequence should be viewed with caution, however, since there may be other smaller errors in it less easily detected by comparison with human and *E.coli* sequences. For instance, figure 22 indicates a two amino acid deletion in the ovine sequence relative to the *E.coli* at positions 434-435. Since this deletion lies immediately adjacent to a boundary of one of the reordered cyanogen bromide fragments, it seems likely that this deletion is an artefact of protein sequencing. Furthermore the five amino acid sequence (SWRRA) aligned opposite amino acids 403-407 in figure 23 may actually be located in the region of this deletion since they exhibit homology with *E.coli* amino acids 430-434. These alterations were not made since they cannot be achieved simply by a rearrangement of the peptide fragments shown in Carne and Walker (1983).

A definitive ovine sequence awaits cloning of the sheep PGD gene or else repetition of the protein sequencing experiments.

The comparison of the adjusted ovine and human

amino acid sequences reveals 46 differing residues out of the 305 compared (fig.23). This represents 85% conservation between the two species (neglecting the gaps between the two sequences).

The availability of the sheep polypeptide sequences permits an estimate of the amount of coding mRNA sequences yet to be cloned assuming the overall lengths of the sheep and human polypeptide sequences to be very similar. The sequence (about 195-210 bases) encoding the residues homologous to the 67 COOH-terminal residues of the sheep polypeptide are undoubtedly within H3. By the same token, the 5' coding sequences of human PGD probably extend at least 70bp (encoding 23 amino acids) beyond the 5' end of H4 and have yet to be cloned.

The coding region of the 2.4kb human PGD mRNA probably covers about 1400 bases:

≥70 bases upstream of H4
\+ 800 bases (H4)
\+ 312 bases (sequenced in H3)
\+ ~200 bases (unsequenced in H3)
~1382 bases

Assuming an amino acid composition in the human unsequenced regions similar to that of the sheep sequence this represents a subunit molecular weight of 52kDal (and about 470 amino acids) which is in exact agreement of with estimates from electrophoresis (Harris and Hopkinson,1976).

The above calculations also suggest that the ORF extends about 500 bp into H3 and therefore only about

250bp of H1 (650bp-see figure 1) consists of open reading frame. This is consistent with the previous evidence suggesting that the PGD rat cDNA, pPGD-1 (which was used to isolate H1) was primarily made up of untranslated sequences (see section 3.1).

CHAPTER 4 DISCUSSION

4.1 Isolation of Human PGD Sequences

Since the initial identification of pPGDH1 (and all the subsequently isolated human sequences) was based on its homology to pPGD-1, it is worth examining the evidence in favour of pPGD-1 representing rat PGD mRNA.

Miksicek and Towle (1982) isolated a PGD antibody from the sera of rabbits repeatedly challenged with purified rat PGD protein. The purity of the PGD protein was assessed by SDS polyacrylamide electrophoresis which demonstrated a single 49kDal band. The specificity of the PGD antibody was demonstrated by electrophoresis of *in vitro* translated mRNA selected by polysome immunoadsorption (Miksicek & Towle,1983). Although the major translation product was identical in size to rat PGD, other smaller translation products were also observed. Thus, some doubt over the specificity of the PGD antibody exists. The subsequent isolation of pPGD-1 was based on:

1. The specificity of the antibody and
2. The induction of PGD mRNA synthesis in rats fed on a high carbohydrate, fat-free diet and administered triiodothyronine. These conditions have been shown to result in up to a sevenfold increase in PGD mRNA levels. However, it should be noted that, under these

conditions, induction of G6PD and malic enzyme mRNA transcription is even greater (Miksicek & Towle,1982).

Thus, although the evidence strongly suggests that pPGD-1 represents part of the rat PGD mRNA, the possibility of this not being the case was considered in the subsequent analysis of the human clones isolated.

The identification of the isolated human cDNA clones as copies of the PGD mRNA is based on four independent lines of evidence:

1. Homology between pPGDH1 and pPGD-1.

The initial evidence that H1 represents part of human PGD mRNA is its homology to the rat cDNA.

2. Northern blot analysis of RNA.

The finding that H4 detects a single human RNA of the same size as rat PGD mRNA (2.4kb) provides indirect support for the authenticity of the human sequences isolated.

3.Physical assignment of H1 and H4.

The localisation of H1 and H4 to a small fragment of the genome within 1p36.13-1pter which is contained within the hybrid DIS2.6, the region to which the PGD enzyme had previously been assigned (Carritt et al.1982) is in itself a strong indication that H1 and H4 represent human PGD mRNA.

4. Linkage analysis.

If H1 represents part of the mRNA which encodes the PGD enzyme subunit, the RFLP detected by H1 and the

enzyme polymorphism would be expected to show very tight linkage. Indeed, no crossovers between the two markers were detected and a maximum lod score of 3.010 for linkage at $\theta=0$ was achieved.

5. Sequence comparison.

The homology between the human cDNA sequences and the previously published PGD sequences of *E.coli* and of sheep strongly support the identification of H3 and H4 as human PGD sequences.

The expectation that the rat PGD cDNA would show homology to human PGD is based on the finding that 'household' enzymes such as PGD are generally highly conserved proteins. Although little evidence is available for the pentose phosphate pathway enzymes, the glycolytic enzymes (which would be expected to have similar functional constraints) are amongst the slowest evolving enzymes known (Fothergill-Gilmore,1986). The only pentose pathway enzyme for which such analysis has been carried out is G6PD and this showed extensive amino acid sequence conservation between fruit-fly, rat and human G6PD (Jeffrey et al.1988). Both the pentose pathway and glycolysis represent ancient pathways present in a large variety of organisms. It would therefore be more probable that rat and human PGD genes are homolgous and derive from the same ancestral gene rather than that they are unrelated genes that encode enzymes which have acquired similar functions by convergent evolution. Furthermore, the evolution of

proteins which operate within pathways is likely to be more constrained than that of other proteins (e.g. the glycolytic enzymes). The existence of many metabolic enzymes including PGD as dimers or tetramers also suggests that they evolve more slowly since at least the region involed in subunit binding needs to be conserved for enzyme activity.

Nevertheless, the initial difficulty in isolating human sequences homologous to the rat partial cDNA suggested a lack of homology between human and rat PGD sequences. The reason for this may have been that pPGD-1 contains only or mainly untranslated sequences which would be expected to show less interspecies homology because the functional constraint is presumably smaller on such sequences. pPGD-1 encompasses about 880bp of the rat PGD mRNA of 2.4kb. Given that the rat PGD enzyme is about 52kDal, it may comprise about 500 amino acids (assuming the average molecular weight of an amino acid to be slightly over 100) and therefore could be encoded by about 1500 nucleotides. Thus, if the 5' untranslated region of the mRNA is short and assuming that the cDNA was primed at the poly(A) tail, there is a quite high probability that pPGD-1 contains little or no coding sequences. Considering this, it was certainly possible that the homology between pPHDH1 and pPGD-1 was spurious and that the two cDNAs did not represent sequences from homologous genes. It was therefore important to accumulate evidence independent of the

homology between the two cDNAs that pPGDH1 represents human PGD.

The similar sizes (2.4kb) of the human and rat mRNAs detected by pPGDH1 and pPGD-1 respectively is suggestive of greater homology extending beyond the cDNAs themselves. This finding is certainly consistent with the similar sizes of rat and human PGD subunits (both 52kDal Bridges et al.1975). As discussed earlier, the PGD gene is likely to be highly conserved and hence the rat and human PGD mRNAs would be expected to be of similar size. Both the rat and human genes for G6PD have been cloned and in this case the protein and RNA sizes do indeed correlate: The human G6PD subunit (56kDal) is encoded by a 2.6kb mRNA whist the rat subunit (59kDal) is encoded by a 2.3kb cDNA. G6PD, however, represents what is at present a unique case in which there is evidence (for human G6PD) that the NH2-terminal amino acids of the subunit are encoded by a separate locus on chromosome 6 (Kanno et al.1989; see later for further discussion). If this example of post-transcriptional processing is confirmed, a predictable relationship between subunit and mRNA size may not always be assumed.

The third line of evidence for the authenticity of the human pPGD clone was the assignment of H1 to the same small region of 1p36 (included in the hybrid DIS2.6) to which PGD has been assigned. The significance of this assignment obviously depends on the

proportion of 1p36 contained within the hybrid, DIS2.6. The estimate presented earlier (section 3.5.1) of 6cM is based on the genetic distance between ENO1 and PND both of which are not present in DIS2.6 (Keats et al.1989). However, the authors themselves emphasize the tentativeness of genetic distance derived from multi-point mapping. This is illustrated by the significant differences between the three maps constructed for 1p (Donis-Keller et al.1987; O'Connell et al.1988; Dracopoli et al.1989- see figure 6). Although there is complete agreement on the order for those loci mapped by more than one group, the genetic distances between them vary considerably. Nevertheless, 6cM is the best available estimate of the maximum genetic distance which the fragment of chromosome 1p in DIS2.6 represents in the genome.

An initial estimate of the physical distance might be $6x10^6$bp using the often-quoted conversion of 1cM to 10^6bp. This is derived from the total genetic size (3300cM) and the total DNA content ($3x10^9$bp) of the genome (Renwick,1971; Davidson, 1976). The value of 3300cM is a sex-averaged genetic length based on chiasma counts in males and adjusted assuming a 40% higher recombination rate in females (Renwick,1971). The difference may actually be as much as 90% (Donis-Keller et al.1987).

$6x10^6$bp may be an over-estimate of the chromosome 1 fragment in DIS2.6 since chiasma frequency maps strongly suggest that the frequency of crossing-over per unit physical length is higher in distal regions of chromosomes and

lower near centromeres. If this is true, it would suggest that a DNA fragment smaller than $6x10^6$ bp within a distal region such as 1p36 may correspond to 6cM. It is difficult, however, to quantify such deviation from the average. Hulten et al. (1982), using chiasma frequency maps estimate the male genetic length of 1p36 to be about 46% of the short arm of chromosome 1 (45.3cM out of 99.0cM) whereas the banded mitotic map suggests that 1p36 represents about 20% of the physical length of 1p. Thus, 1p36 may experience recombination roughly twice as often as the average for 1p such that a genetic distance of 6cM in 1p36 may be equivalent to $3x10^6$ bp.

The above calculation suggests that the assignment of H1 (and H4) to the region of chromosome 1 included in DIS2.6 is highly significant since this amount of DNA represents, at most, about 0.1% of the genome. It is very unlikely that a cDNA isolated on the basis of homology to a rat PGD cDNA should detect sequences in this region of the genome by chance.

The tight linkage between the RFLP detected by H1 and the PGD enzyme polymorphism provides strong support, independent of the somatic cell hybrid data, that H1 is linked to PGD. A lod score of 3.0 means that linkage between the markers at $\Theta=0$ is 1000 times more likely than this result being obtained by chance and that the markers are actually unlinked. In fact the prior probability of two randomly-chosen autosomal

markers being linked is 1/50 (Morton,1955) such that a lod score of 3 suggests only odds of only 20:1 in favour. This is generally considered to be sufficient for establishment of linkage.

If on the other hand, the linkage is considered in the light of the prior assignment of both markers to a small region (equivalent to about 6cM) within 1p36.13-1pter, then the null hypothesis of no linkage used in the calculation of lod scores is no longer appropriate. As described earlier, a calculation of the log of odds of $\Theta=0$ versus $\Theta=0.06$ in an analagous way to the calculation of lod scores shows that the data suggests that linkage at $\Theta=0$ is roughly twice as likely as linkage at $\Theta=0.06$. In order to improve this likelihood to a significant level (20:1), roughly ten times as much data would need to be collected. This illustrates the effort required in order to accurately determine small genetic distances in man. In the present context of determining the authenticity of H1, this was considered to be unnecessary given the independent evidence presented previously.

The homology between the human sequences and the *E.coli* PGD amino acid sequence of 50% is obviously highly significant. For example, adenesine phosphoribosyltransferase shows only a 40% identity between *E.coli* and human. On the other hand, *E.coli* and human GMP reductase (also a NADP-dependent oxidoreductase) share 66% homology. Thus, the 50% homology between the

E.Coli and human PGD sequences certainly falls within the expected range for a ubiquitous enzyme such as PGD.

The 46 residue differences between the 305 amino acids compared in sheep and human PGD (15% non-homology; see section 3.13) can be expressed as an average rate of amino acid substitution for PGD. This represents 16.5 PAMs (accepted point mutations per 100 residues). This figure incorporates a correction to take account of superimposed substitutions (Dayhoff,1976). Assuming 75 million years since the divergence of sheep and man (Dayhoff,1976) and equal rates of evolution in the two lineages, this gives a value of about 11 PAMs per 100 million years. This compares with rates of 4-6PAMs/100 million years for most glycolytic enzymes and much higher values for proteins under little functional constraint such as fibrinopeptide which evolves more rapidly at about 90PAMs/100 million years).

Indeed, a pentose pathway enzyme such as PGD might be expected to show similar rates of evolution as the glycolytic enzymes. The evolution of an enzyme that is part of a multi-step pathway is constrained by the necessity to recognize similar ligands and substrates as the enzymes catalyzing the previous an dsubsequent steps of the pathway. Thus, in general, enzymes of a single pathway are likely to evolve at similar rates which will be slower than those for enzymes which act in a more independent manner (e.g. ribonuclease A).

This has indeed been shown to be the case for many glycolytic enzymes (Fothergill-Gilmore,1986). The slightly higher rate of evolution calculated for PGD may reflect the less central role of the pentose pathway compared to the glyolytic sequence in most cells. However, a brief examination of the homologous amino acid residues in sheep and human PGD (see fig.20) shows that this average value for the rate of evolution of PGD is rather misleading since the differences are generally restricted to certain regions (e.g. between residues 73 and 89 in the ovine sequence) which separate long stretches of almost completely homologous sequences. For example, an 87 amino acid sequence in sheep (311-397) is different from the human sequence at only 3 residues. This suggests that this region has evolved at less than 2 PAMs/100 million years. Such strong conservation indicates that this region is of crucial importance to the activity of the enzyme. This is supported by the fact that the *E.coli* sequence in this region is highly homologous to the human and ovine sequences.

Indeed, the average rate of evolution of PGD derived from the comparison of *E.coli* and human sequences (150 different residues out of 303 compared) is less than 3 PAMs/100 million yrs. This is significantly lower than the value derived from the human-sheep comparison. However, it is based on an estimate of 1500 million years for the time elapsed since the

divergence of prokaryotes and eukaryotes which is, at best, a rough approximation (Fothergill-Gilmore,1986). The value of 11 PAM/100 residues/100 million years obtained from the sheep-human comparison is likely to be more accurate.

The evolution of metabolic pathways themselves can be addressed by examining the homologies between the enzymes that make up such pathways. Most theories of pathway evolution assume that these enzymes have arisen through divergent evolution of enzymes from a smaller number of progenitors. Thus, enzymes that catalyze similar reactions or perhaps those with similar ligand requirements would be ancestrally related. Another possibility is that the enzymes that catalyze sequential steps of a pathway may have evolved from a single common gene. Alternatively, it may be that the similar domains of different enzymes are the result of convergent evolution of unrelated enzymes and that pathways develop from the chance association of independently-evolving enzymes.

The glycolytic enzymes have been closely examined in this context and although there are cases of probable divergent evolution (e.g. glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase), the pathway seems to have arisen from the random association of unrelated enzymes (Fothergill-Gilmore,1986).

In the light of this, the human PGD sequences were compared with those of G6PD from several species

(human, opossum and Drosophila). G6PD catalyzes the dehydrogenation of glucose-6-phosphate as the initial rate-limiting step of the pentose pathway. Like PGD, it is a NADP-dependent oxidoreductase which is induced by triiodothyronine. However, no homology was detected at either the DNA or amino acid level suggesting that as in the case of the glycolysis the oxidative part of the pentose pathway has arisen from the chance association between G6PD and PGD. Another enzyme which may be related to PGD is 'malic' enzyme (NADP-dependent malate dehydrogenase) which, like PGD and G6PD, generates NADPH reducing power for lipogenesis and is also induced in the liver by triioidothyronine (Hutchison & Holten,1978). The rat gene encoding this enzyme has been cloned (Morioka et al.1989) and the mRNA sequence was compared with that of PGD. As before no sequence homology was detected at the DNA or amino acid level. In this context, it is interesting to note, however, that one of the minor signal peaks observed when tritiated H1 was hybridized to metaphase chromosomes was at 6p12-6q12, a region which includes the locus for human cytoplasmic malic enzyme (6q12; Meera Khan et al.1984).

Nevertheless, sequence analysis and somatic cell hybrid data suggest that PGD, G6PD and malic enzyme do not share common ancestry and it is possible that their analogous catalytic activities is the result of secondary or tertiary structure similarities not reflected

in the amino acid sequences of these proteins.

The availability of PGD amino acid sequence obtained directly from analysis of the protein (sheep) and also cDNA-derived PGD amino acid sequence (*E.Coli* and human) provides the opportunity to detect possible post-transcriptional processing of PGD. In particular, by comparing amino acid and mRNA sequences it is possible to check for the possible occurence 'ribosome hopping' or transpeptidation. These phenomena have been proposed as two possible explanations for the disparity between G6PD protein and mRNA sequence. Kanno et al.(1989) showed that the major G6PD subunit in human red cells is a chimeric protein in which the 479 carboxy-terminal end residues are encoded by G6PD mRNA but the 55 amino-terminal are encoded by an unlinked locus on chromosome 6 (later shown to encode GMP reductase; Henikoff & Smith,1989). Since no mRNA encoding all 534 amino acids was detected, it is suggested that the chimeric protein is the consequence of either the translocation of the ribosome from the GMP reductase mRNA to the G6PD mRNA or the cleavage and joining of the two polypeptides. The existence of a common 15 base sequence in the 5' leader sequences of both mRNAs suggests the possibility of alignment of the two mRNAs prior to translation to allow 'ribosome hopping'. It may also be significant that the first residue encoded by the G6PD mRNA (at position 56 in the polypeptide) is methionine.

However, no other occurrence of this phenomenom has been reported and its possible function remains obscure since a minor fraction of the G6PD subunits in the human red cell were found to be solely encoded by G6PD mRNA. Also, G6PD clearly segregates with the X chromosome (the G6PD locus lies at Xq28; Pai et al.1980) in somatic cell hybrids. These facts and the failure by Kanno et al. (1989) to detect GMP reductase activity in rat and mouse red cells show that this novel form of polypeptide processing, if it exists, is not essential for G6PD activity.

In the case of PGD, the extensive homology between the entire length of the sheep peptide sequence and the *E.coli* or human cDNA sequences suggests that ribosome hopping or transpeptidation is not involved in the expression of PGD in lamb liver. No amino sequence of PGD obtained from red cells has been published. Since there is no suggestion that the chimeric G6PD protein is ubiquitous amongst mammalian species (it is probably absent in rat and mouse red cells; Kanno et al.1989), it would of interest to sequence the human red cell PGD protein. In this context it is worth remembering the crucial role played by the pentose pathway in the protection of the red cell from its highly oxidative enviroment. One could envisage that the synthesis of chimeric G6PD (and PGD) with different characteristics to their 'normal' counterparts might be a simple way of altering pentose pathway function in a cell-specific

manner. It is also noteworthy that G6PD is the rate-limiting enzyme of the pentose pathway and alterations of this enzyme may well occur without similar changes in PGD.

Determination of sequence 5' of the transcription start would provide useful information concerning the regulation of PGD. The 5' regulatory sequences of the G6PD gene have been determined (Martini et el.1986) and a comparison of the two may provide clues to their common regulation as well as to the differences in their regulation. G6PD possesses a non-canonical TATA box, nine GGGCGG (or its complement,CCCGCC) hexa-nucleotides and no CAAT element. The AT-rich TATA box (Breathnach & Chambon,1981) influences the site of initiation (Benoist & Chambon,1981; Mathis & Chambon,1981) and the GGGCGGG sequence is thought to interact with the transcription factor Sp1 (Kadonga et al.1986). The CAAT element is often found in eukaryotic genes 70 to 80 nucleotides upstream of the start site and is known to affect the level of transcription (Efstradiatis et al.1980). It would be of interest to investigate the extent of the similarity and the differences between PGD and G6PD with respect to the elements described above.

Furthermore, one might expect the existence of T3-resposive element in the PGD promoter region. This is a palindromic sequence similar in structure to the reti-noic acid and glucorticoid-responsive elements (RRE &

GRE; Beato,1989) which has been shown to be involved in the induction of genes by triiodothyronine (T3) receptor (Glass et al.1987; Koenig et al.1987).

4.2 PGD-Related Sequences

Hybridisation of H1 and H4 to human genomic DNA showed that although both cDNAs lack internal EcoRI sites, they both detect several EcoRI fragments. This could be explained by the hybridisation of the cDNA to exons which, in the genome, are separated by introns or it could be due to the presence of several unlinked PGD-related sequences.

Somatic cell hybrid analysis showed that all three H1-hybridizing EcoRI fragments were on different chromosomes thereby demonstrating the existence of at least two unlinked PGD-like loci. The most obvious possibility is that these represent PGD pseudogenes. i.e. transcriptionally defective related sequences. An alternative possibility is that the PGD cDNA hybridizes to related genes such as those encoding glucose-6-phosphate dehydrogenase or malic enzyme both of which are NADP-dependent dehydrogenases. In the light of the hybridisation of H4 to a single RNA on Northern blots, this possibility seems unlikely.

The large majority of pseudogenes are processed such that they show homology to the mature mRNA but lack the intervening sequences of the functional gene. Most processed pseudogenes contain a poly(A) sequence at their 3' end and the homology shared with the functional gene does not extend beyond the mRNA cap nucleotide. Thus, it seems probable that these pseudo-

genes derive from mature mRNAs either by direct integration of the RNA or by reverse transcription of the RNA into a cDNA followed by integration (see Vanin,1985; Wagner,1986 for reviews).

The majority of processed pseudogenes are homologous to genes which are ubiquitously expressed. Indeed, if they are derived from mRNA, their expression in germ cells is a prerequisite for their formation and constitutive presence in the genome. The existence of processed pseudogenes homologous to genes not expressed in the germ-line (e.g. globins) is thought to be due to aberrant transcription (Vanin,1985). PGD is expressed in both ovary and testis germ cells (Glock & McLean,1954) and the existence of processed pseudogenes of PGD is therefore not unexpected.

The accepted mechanism of the formation of processed pseudogenes predicts that their chromosomal location will not be related to that of the parent gene and ,indeed, this has been found to be the case (Vanin,1985). For example, some of the 40-odd glyceraldehyde-3-phosphate pseudogenes have been assigned to 12 different chromosomes (Benham & Povey,1987).

The 1.25kb EcoRI fragment detected by H1 in genomic DNA segregated unambiguously with chromosome 18 in a panel of somatic cell hybrids and this assignment was refined by *in situ* hybridisation 18q21.1-12.1. This is certainly consistent with this being the locus of a PGD processed pseudogene.

Processed pseudogenes can often be distinguished from functional genes if a cDNA probe hybridizes across introns which possess one or more sites for a restriction enzyme which does not cleave within the probe itself. If genomic DNA is digested with such a restriction enzyme the cDNA will detect several fragments at the gene locus, but only one at each pseudogene. Hybridisation of H4 to EcoRI-cut hybrid DNA showed precisely this: Three bands segregate with chromosome 1p whilst the remaining two fragments segregate independently. Analysis with other restriction enzymes was consistent with these results.

It may be significant that the size of the smallest EcoRI band (0.9kb) detected by H1 is similar to that of the two independently-isolated PGD cDNA clones H2 and H3. If, as the evidence presented earlier suggests, the EcoRI ends of the two cDNAs are natural, this would imply that a processed pseudogene of PGD which has retained sufficient homology to the gene would contain a 0.9kb H1-hybridizing EcoRI fragment. However, the relatively weak hybridisation of H1 to the 0.9kb EcoRI genomic fragment is suggestive of weak homology. It is of course possible that the sizes of the EcoRI fragment and the cDNA clones is coincidental.

It is interesting that the existence of PGD processed pseudogenes *per se* lends indirect support to the supposition that H2 and H3 are bounded by natural EcoRI sites: Nearly all processed pseudogenes have been

found to contain a poly(A) tract at their 3' ends. Indeed, the poly(A) tails of mRNAs are thought to be intimately involved in the formation of pseudogenes (Moos & Gallwitz,1982; Vanin,1985). All those processed genes without poly(A) tracts have until now been found to be homologous to genes with tissue-specific expression such as mouse α_3-globin. These are likely to have been formed by different mechanisms (Goff et al.1980; Vanin, 1985). Since PGD is expressed in the germ-line, the formation of PGD pseudogenes is likely to involve a poly(A) tail which obviously implies that PGD mRNA is polyadenylated. If this is the case, the absence of a poly(A) tract in three independently cloned cDNAs suggests that they may have been shortened during cloning presumably by EcoRI digestion.

In contrast with the other two EcoRI bands, the 0.9kb EcoRI fragment detected by H1 did not segregate with any single chromosome in a panel of 18 hybrids. This result may reflect one or more of the following:

1. The 0.9kb band represents a single locus, but some hybrids are incorrectly typed with respect to their human chromosome complement.
2. The 0.9kb band represents a single locus, but some hybrids are incorrectly typed with respect to this band. This possibility should be considered since the hybridisation of H1 to this fragment was considerably weaker than to either of the other two EcoRI fragments.
3. The 0.9kb band represents several loci which share

sufficient homology to produce identically-sized H1-hybridizing EcoRI fragments. However, if this band reflects the existence of several PGD pseudogenes sufficiently homologous to have retained the same-sized H1-hybridizing EcoRI fragment, the expectation would be that the intensity of the 0.9kb EcoRI band would be greater than that of the 1.25kb band.

In situ hybridisation of H1 to metaphase chromosomes yielded little additional information to distinguish between these possibilities. A major peak of signal at 1p36 and a single minor peak at 18q were detected. The third peak around the centomere of 6q was weak and not consistent with the somatic cell hybrid analysis. the generally high level of background may indicate the presence of several loci.

Analysis of H4-hybridising EcoRI fragments in a panel of somatic cell hybrids showed that the segregation of the 3kb band was consistent with that of chromosome 18. However, when hybridized to metaphase chromosomes, H4 did not detect any sequences on chromosome 18.

The presence of a co-migrating hamster band precluded the assignment of the 0.75kb, H4-hybridizing EcoRI fragment.

It should be noted that since H1 and H4 share no homology, the loci detected by these two cDNAs need not be the same. The number of PGD-like sequences in the genome is thus unknown. This issue might be resolved by

the hybridisation of H3 and H4 to a panel of somatic cell hybrids digested with a variety of restriction enzymes. However, this was considered to be of marginal interest in the context of this project.

4.3 PGD as a Genetic and Physical Marker

The detection of an RFLP by H1 has obviously increased the genetic information available at the *PGD* locus. *PGD* was one of the first loci to be assigned to chromosome 1p (by establishment of linkage between the common PGD enzyme polymorphism and the Rhesus locus; Weitkamp et al.1971). It was therefore used as a marker in the construction of the earlier genetic and physical maps of chromosome 1 (e.g. Hamerton & Cook,1974).

However, the advent of DNA markers and the higher heterozygosity associated with RFLPs has severely limited the relative usefulness of enzyme markers. This is compounded by the possibility of physically assigning DNA probes by *in situ* hybridisation (and more easily than enzyme markers by somatic cell hybrid analysis) and more recently, by pulse-field gel electrophoresis (PFGE).

The cloning of the *PGD* locus will facilitate the incorporation of this locus into more recent genetic and physical maps of chromosome 1p using this techniques. The BamHI RFLP detected by H1 with an estimated PIC (polymorphic information content) of 0.30 is a significant improvement on the most common PGD polymorphism (PIC=0.04 in a London population; based on the data from Parr,1966). Indeed, this RFLP is presently being mapped relative to other 1p markers in another laboratory (N.Dracopoli) in an effort to isolate a

locus involved in cutaneous malignant melanoma (CMM) which has been assigned to 1p36 (Bale et al.1989). Since tight linkage between this RFLP and the PGD enzyme polymorphism has been established, the linkage data obtained from use of this RFLP can justifiably be incorporated with the existing PGD enzyme polymorphism data.

The genetic map of chromosome 1p has progressed rapidly during the course of this project (Donis-Keller et al.1987,1989; Dracopoli et al.1988; O'Connell et al.1989) and there are now 9 polymorphic cloned markers (including 6 gene loci) which have been assigned to chromosome 1p36 and many more which have been assigned to larger segments of chromosome 1p which include 1p36 (Kidd et al.1989).

Given this number of markers in a region of the genome which includes roughly $3x10^7$bp (20% of chromosome 1 which includes about 5% of the $3x10^9$bp genome; Paris Conference,1971), it is feasible to begin more accurate physical mapping of segments of this chromosomal band by PFGE using these markers.

Physical maps of some regions with a high density of markers have already been constructed. For example, a preliminary map of a distal region of the long arm of the X chromosome (Xq28) spanning 1700kb and seven cloned loci has been constructed by PFGE (Arveiler et al.1989).

A physical map of the X-Y pseudoautosomal region

has been constructed by Brown (1988) and Petit et al.(1988). The region has been estimated to span 2600kb and the recombination fraction across it is 50% (since this region is defined as that which undergoes an obligate recombination event during male meiosis). Thus , 1cM corresponds to roughly 50-60kb (Petit et al.1988). This represents an increase of about 20-fold on the average for the genome (1cM-10^6bp). Both Brown (1988) and Petit et al.(1988) found that the relationship between genetic and physical distance was uniform along the length of this region.

The possible tight physical linkage between PGD and ENO1 on a 350kb NarI fragment allows a preliminary analysis of relationship between genetic and physical distance in this region of the chromosome. The present estimate of the genetic distance between ENO1 and PGD is 1.5cM (Morton,1988; Keats et al.1989). This suggests that 1cM is equivalent to a physical distance of, at most, 230kb in this region of the genome. This represents an increase of almost five-fold on the average ratio of physical to genetic distance in the genome (10^6:1cM). Rough estimates of the physical length of the 1p36 (about $3x10^7$bp-see above) and its genetic length (45cM; Hulten et al.1982) indicate an average ratio of approximately 1cM to 600kb for this region. The estimate derived from the ENO1-PGD linkage is of the same order of magnitude.

The observed synteny of PGD and enolase in 12

different primates and 7 other mammals (Lalley et al.1989) is certainly consistent with their tight physical linkage. These loci are even syntenic in platypus, a prototherian mammal which diverged from placental and marsupial mammals 150-170 million years ago (Lalley et al.1989). Rabbit is the only mammal in which the loci encoding enolase and PGD activities have been assigned to separate chromosomes (Lalley et al.1989).

Conclusive demonstration of physical linkage between pPGDH4 and ENO1 is hampered by the detection of at least one other locus by H4. This problem may be resolved by assigning PFGE bands by somatic cell hybrid analysis bearing in mind the fact that most rare-cutting restriction enzymes used are methylation sensitive and that the methylation pattern of human DNA in a rodent cell may well be different to that in a human cell. Thus, the same DNA regions in hybrids and human cells may give rise to very different fragments. In fact, the different fragment sizes may aid the establishment of linkage.

Confirmation of the ENO1-PGD linkage may be simplified by the use of a PGD locus-specific probe rather than pPGDH4. The most obvious choice would be an intronic probe which would not detect processed pseudogenes. P1.4 (see figure 16) may also be an appropriate probe since it contains at least part of the CpG island. Pseudogenes of PGD are unlikely to have

retained sequence homology with this region of the transcribed gene.

An advantage of gene probes associated with CpG islands is that their location with respect to the sites for many of the restriction enzymes used for PFGE analysis is known. Therefore unambiguous maps using these enzymes can be constructed. Indeed, the use of two probes either side of the CpG island should permit the construction of a physical map extending 100-1000kb depending on the CpG island density in this region.

Extension of physical mapping to other nearby markers (e.g.D1Z2 or D1S43; Buroker et al.1987; Donis-Keller et al.1987) would provide more accurate estimates of the genetic to physical distance ratio in this region. The frequency of recombination in the region surrounding D1Z2, a VNTR (variable number tandem repeat) locus would be of particular interest given the possible involvement of sequences of this type in recombination. The sequence similarity between VNTRs and the Chi recombination sequence of bacteriophage λ and *E.coli* led Jeffrey et al.(1985) to propose that these sequences are eukaryotic recombination signals. However, there is evidence from VNTR sequences that do not seem to be hot-spots of crossing-over but which nevertheless are located in regions of high recombination rates (Higgs et al.1988) that VNTRs may simply be the result of frequent recombination.

Whether VNTRs are a cause or a consequence of high

recombination frequency, it would be interesting to discover if crossover events are clustered in 1p36 or if this region is similar to the pseudoautosomal region in which no evidence for clustering has been found (Petit et al.1988; Brown,1988).

APPENDIX A PGD^B ALLELE FREQUENCIES

POPULATION	GENE FREQUENCY PGD^B	REFERENCE.
European		
London,U.K.	0.021	Parr,1966
Chicago,U.S.A.	0.039	Bowman et al.,1966
Chicago,U.S.A.	0.009	Dern et al.,1966
Cape Town,S.A.	0.033	Gordon et al.,1967
Jo'burg,S.A.	0.013	Jenkins et al.,1971
Buffalo,U.S.A.	0.024	Davidson et al.,1967
Greece	0.039	Carter et al.,1968
Seattle,U.S.A.	0.019	Giblett,1969
Iceland	0.022	Tills;M.Phil,1969[1]
Ireland	0.014	as above
Skolt Lapps	0.024	as above
Sevettijärvi Lapps	0.032	Eriksson; in Tills et al.,1970 & 1971
Lapps (N.Sweden)	0.130	Beckman et al.,1971
Asiatic		
Malaysia		
Malay	0.032	Eng & Welch,1972
Chinese	0.031	as above
Indian	0.014	as above
Aboriginal(Misc.)	0.055	as above
Semai	0.037	as above
Temiar	0.064	as above
Malay,Cape Town	0.025	Gordon et al.,1967
China	0.073	Shih et al.,1968[1]
Taiwan	0.059	as above
Hokkaido,Japan	0.069	Giblett,1969
Ainu,Japan	0.071	Tills et al.,1970/71
Thailand	0.070	Giblett,1969
Bhutan	**0.231**	Fitch & Parr,unpub.; in Carter et al.,1968 Mourant et al.,1968[1]
Seattle,U.S.A. (Mixed)	0.048	Giblett,1969[1]
Papua	0.087	Tills et al.,1970/71
Nepal	0.096	as above

POPULATION	GENE FREQUENCY PGD^B	REFERENCE.
Middle East		
Iran(Moslems)	0.028	Bowman et al.,1967[1]
Israel		
Kurdish Jews	0.038	Tills,M.Phil,1969[1]
Yemenite Jews	0.092	as above
Habbanite Jews	0.074	Fitch & Parr,unpub.; in Carter et al.,1968
Kurds	0.047	Tills et al.,1970/71
Arab		
South Arabia	0.110	Tills,M.Phil,1969[1]
Jordan	0.065	as above
Palastinian	0.075	as above
Negroid		
Mozambique	0.091	Giblett,1969
Zambia	0.059	Tills;M.Phil,1969[1]
Nigeria	0.060	Parr & Fitch,1966
Uganda	0.055	as above
South Africa		
Xhosa(Cape Town)	**0.139**	Gordon et al.,1967
Zulu(Durban)	0.104	Kirk et al.,1971[2]
All below from Jenkins & Nurse,1974		
Zulu/Tonga	0.033	
Swazi	0.061	
Sotho (South)	0.085	
Sotho (West)	0.060	
Sotho (North)	0.071	
Botswana		
Herero	**0.000**	
Namibia (Ambo)		
Kuanyama	0.024	
Ndonga	0.042	
Kuambi	0.062	
Ngandjera	0.059	
Mbalatu	0.028	
Kualuthi	0.079	
Nkolonkadi	0.031	
Namibia (Kavango)		
Kuangali/Mbunza	0.050	
Sambyu	0.019	
Botswana (Kavango)		
Dciriku	0.047	
Mbukushu	0.048	
Chicago,U.S.A.	0.037	Bowman et al.,1966
Chicago,U.S.A.	0.030	Dern et al.,1966
Buffalo,U.S.A.	0.036	Davidson et al.,1967
Seattle,U.S.A.	0.055	
Uganda and Kenya	0.070	Tills et al.,1970/71

POPULATION	GENE FREQUENCY PGD^B	REFERENCE.
Hadza,	0.006	Tills;M.Phil,1969[1]
Amhara,Ethiopia		Carter et al.,1968
Adi-Arkai(1500m)	**0.053**	Mourant et al.,1968[1]
Debarech(3000m)	**0.135**	& Harrison et al.,1969
Beja,	0.155	Carter et al.,1968
		Hassan et al.,1968[1]
Nyanturu,	0.030	Tills et al.,1970/71
Sandawe,	0.069	as above
Nigeria	0.078	as above
American Indians		
Venezuela		
Yanomama	**0.002**	Arends et al.,1967
Mexico		
Lacandon/Gp1	**0.000**	Bowman et al.,1966
Lacandon/Gp2	0.213[3]	as above
Yucatan	**0.000**	as above
Brazil	**0.000**	Tashian et al.,1967
Dominica		
Caribs	**0.000**	Glasgow et al.[1]
		Tills,M.Phil,1969[1]

POPULATION	GENE FREQUENCY PGD^B	REFERENCE.
Australian Aborigines		
Arnhem Land		
Elcho Land	0.040	Blake and Kirk,1969
Other	0.057	as above
Central Australia	0.041	as above

[1] Quoted in Tills et al.,1970 & 1971
[2] Quoted in Jenkins & Nurse,1974
[3] Estimated from PGD^B PGD^B frequency (Tills et al.,1970)

APPENDIX B ELECTROPHORETIC VARIANTS OF PGD ENZYME

VARIANT	PRESUMED GENOTYPE	SLOW/ FAST[1]	SOURCE
Normal	*PGD*^{A} *PGD*^{A}	-	Fildes & Parr,1966
Common (Plaistow)	*PGD*^{A} *PGD*^{B}	S	as above
Canning	*PGD*^{B} *PGD*^{B}	S	Parr,1966
Richmond	*PGD*^{A} *PGD*^{R}	F	Parr,1966; Davidson,1967
Hackney	*PGD*^{A} *PGD*^{H}	S	Parr,1966
Friendship[2]	*PGD*^{A} *PGD*^{F}	F	Davidson,1967
Elcho	*PGD*^{A} *PGD*^{Elcho}	F	Blake & Kirk,1969
Freiburg	*PGD*^{A} *PGD*^{Freiburg}	F	Tariverdian et al.1970
Singapore	*PGD*^{A} *PGD*^{Singapore}	S	Blake et al.1973
Wancoat	*PGD*^{A} *PGD*^{Wancoat}	S	as above
Canberra	*PGD*^{A} *PGD*^{Canberra}	S	as above
Kadar	*PGD*^{A} *PGD*^{Kadar}	F	as above
Caspian	*PGD*^{A} *PGD*^{Caspian}	F	as above
Bombay	*PGD*^{A} *PGD*^{Bombay}	F	as above
Thai[3]	*PGD*^{A} *PGD*^{Thai}	F	Tuchinda et al.1968 Eng et al.1972
Natal[4]	*PGD*^{A} *PGD*^{Natal}	F	Blake et al.1974
Oshakati	*PGD*^{A} *PGD*^{S}	S	Jenkins & Nurse,1974
Medi-terranean	*PGD*^{A} *PGD*^{Med}	S	Corbo et el.1981 Nevo et el.1989

[1] Slow and fast refer to the mobility of the enzyme on a starch gel relative to the *PGD*^{A} *PGD*^{A} dimer.

[2] The Friendship variant electrophoretic pattern is only detected if NADP is present in the gel. Otherwise it is indistinguishable from that of *PGD*^{A} *PGD*^{A} (Davidson,1967).

[3] Tuchinda et al.(1968) and Blake et al.(1974) both suggested that the Thai variant may well be identical to the Richmond variant whilst Jenkins and Nurse (1974) suggested it was the same as the Elcho variant. Many of these variants have not been studied side by side so it may be that others are also identical.

[4] The Natal variant electrophoretic pattern is altered if the enzyme is pre-treated with reagents which react with sulphydryl groups (Blake et al.1974).

REFERENCES

Ajmar F., Scharrer B., Hashimoto F. and Carson P. (1968). Interrelation of stromal NAD(P)ase and human erythrocytic 6-phosphogluconate dehydrogenase. Proc. Natl. Acad. Sci. USA **59**; 538.

Arora K.K., Cortis P., Bleakley P.A., Williams J.F. (1985). Identification and measurement of D-glycero D-iodo octulose 1,8-biphosphate: D-alto heptulose 7-phosphotransferase enzyme in tissues with L-type pentose phosphate pathway. Int. J. Biochem. **17**; 1329.

Arveiler B., Vincent A.and Mandel J.L. (1989). Toward aphysical map of the Xq28 region in man: Linking color vision, G6PD and coagulation factor VIII gene to an X-Y homolgy region. Genomics **4**; 460.

Avery O.T., MacLeod C.M. and McCarty M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J. Exp. Med. **79**; 137.

Aviv H. and Leder P. (1972). Purification of biologically active mRNA by chromatography on oligo-thymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**; 1408.

Bale S.J., Dracopoli N.C., Tucker M.A., Clark W.H., Fraser M.C. Green P., Donis-Keller H., Housman D.E. and Greene M.H. (1989). Mapping the gene for hereditary cutaneous malignant melanoma-dysplatic nevus to chromosome 1p. N, Engl. J. Med. **320**; 1367.

Barker D., Green P., Knowlten R., Schumm J., Lander E., Oliphant A., Willard H., Akots G., Brown V., Gravius T., Helms C.,Nelson C., Parker C.,Redikker K., Rising M., Watt D., Weissenbach B. and Donis-Keller H. (1987). Genetic linkage map of human chromosome 7 with 63 DNA markers. Proc. Natl. Acad. Sci. USA **84**; 8006

Bell J. and Haldane J.B.S. (1937) The linkage between the genes coding for colour-blindness and hemophilia in man. Proc. Roy. Soc. B **123**; 119.

Benham F.J., Hodgkinson S., and Davies K.E. (1984). A glyceraldehyde-3-phosphate dehydgogenase pseudogene on the short arm of the human X chromosome defines a multigene family. EMBO J. **3**; 2635.

Benham F.J. and Povey S. (1987). Members of the human glyceraldehyde-3-phosphate dehydrogenase-related gene family are widely dispersed in the human genome. Cytogenet. Cell Genet. **46**; 579.

Benton W.D. and Davis R.W. (1977). Screening gt recombinant clones by hybridisation to single plaques. Science **196**; 180.

Bird A.P. (1986). CpG-rich islands and the fuction of DNA methylation. Nature **312**; 209.

Bird A.P. (1987). CpG islands as gene markers in thje vertebrate nucleus. Trend. Genet. **3**; 342.

Birnhoim H.C. and Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl.Acid Res.**7**; 1513.

Blake N.M., Saha N., McDermid E.M. and Kirk R.L. (1974). Additional electrophoretic variants of 6-phosphogluconate dehydrogenase. Humangenetik **21**; 347.

Boehnke M., Arnheim N., Li H. and Collins F.S. (1989). Fine-structure genetic mapping of human chromosomes using the polymerase chain reaction on single sperm: Experimental design considerations. Am. J. Hum. Genet. **45**; 21 (1989).

Borck K., Beggs J.D., Brammer W.J., Hopkins A.S. and Murray N.E. (1976). The construction *in vitro* of transducing derivatives of phage lambda. Mol. Gen. Genet. **146**; 199.

Botstein D., White R.L., Skolnick M. and Davis R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Gen. **32**; 314.

Bowman J.E., Carson P.E. and Frischer H. (1966) Genetics of starch-gel electrophoretic variants of human 6-PGD (6-phosphogluconate dehydrogenase): Population and family studies in the Unites States and in Mexico. Nature **210**; 811.

Boyer S.H., Porter I.H., Weibacher R.G. (1962). Electrophoretic heterogeneity of G6PD and its relationship to enzyme deficiency in man. Proc. Natl. Acad. Sci. **48**; 1868.

Bradley D.G., Farrar G.J., Sharp E.M., Kenna P., Humphries M.M., McConnell D.J., Daiger S.P., McWilliam P. and Humphries P. (1989) Autosomal Dominant Retinitis Pigmentosa: Exclusion of the gene from the short arm of chromosome 1 including the region surrounding the Rhesus locus. Am. J. Hum. Genet. **44**; 570.

Brewer G.J., Tavlov A.R. and Alving A.S. (1962). The methemoglobin reduction test for primaquine sensitivity of erythrocytes. J. Am. Med. Ass. **180**; 386.

Brewer G.J. and Dern R.J. (1964). A new inherited enzymatic deficiency of human erythrocytes: 6-phosphogluconate dehydrogenase deficiency. Am. J. Hum. Genet. **16**; 472.

Bridges R.B., Palumbo M.P., Wittenburger C.L. (1975). Purification and properties of an NADP-specific 6-phosphogluconate dehydrogenase from *Streptococcus faecalis*. J. Biol. Chem. **250**; 6093.

Brodie A.F. and Lippmann F. (1955). Identification of a Gluconolactoase. J. Biol. Chem **212**; 677.

Brown W.R.A. (1988). A physical map of the human pseudoautosomal region. EMBO J. **7**; 2377.

Bruns G.A.P. and Sherman S.L. (1989). Report of the committee on the gentic constitution of chromosome 1. Cytogenet. Cell Genet. **51**; 67.

Buroker N., Bestwick R., Haight G. Magenis R.E. and Litt M. (1987). A hypervariable repeated sequence on human chromosome 1p36. Hum Genet. **77**; 175.

Carne A. and Walker J.E. (1983). Amino acid sequence of ovine 6-phosphogluconate dehydrogenase. J. Biol. Chem. **258**; 12895.

Carritt B. (1980). Mapping structural and other genes using somatic cell hybridisation. in "Tissue Culture in Medical Research II" Richards R.J. and Rajan K.T. (eds). Pergamon Press. Oxford and New York.

Carritt B., King J. and Welch H.M. (1982). Gene order and localization of enzyme loci on the short arm of chromosome 1. Ann. Hum. Genet. **46**; 329.

Carritt B. and Povey S. (1979). Regional assignment of the loci *AK3, ACONS and ASS* on human chromosome 9. Cytogenet. Cell Genet. **23**; 171.

Carson P.E., Ajmar F., Hashimoto F. and Bowman J.E. (1966). Electrophoretic demonstration of stromal effects on haemolysate G6PD and 6PGD. Nature **210**; 813.

Carson P.E., Flanagan C.L., Ickes C. and Alving A.S. (1956). Enzymatic deficiency in primaquine-sensitive erythrocytes. Science **124**; 484.

Chirgwin J.M., Przybyla A.E., MacDonald R.J. and Rutter W.J. (1979). Isolation of biologically active ribonucleic acid from sources enrichred in ribonuclease. Biochemistry **18**; 5294.

Connor-Johnson B.C. and Sassoon H.F.(1967). Studies on the induction of liver glucose-6-phosphate dehydrogenase in the rat. Enz. Reg. 5; 93.

Cori O. and Lippmann F. (1952). The primary oxidation product of enzymatic glucose-6-phosphate oxidation. J. Biol. Chem. **194**; 417.

Damm K., Thompson C.C. and Evans R.M. (1989). Protein encoded by v-*erbA* functions as a thyroid hormone receptor antagonist. Nature **339**; 593.

Dayhoff M. (1976). Atlas of Protein Sequence and Structure. National Biomedical Research Foundation. Washington D.C.

DeMoss R.D., Bard R.C. and Gunsalus I.C.(1951). The mechanism of the heterolactic fermentation: A new route of ethanol fermentation. J. Bact. **62**; 499.

Dern R.J., Brewer G.J., Tashian R.E. and Shows T.B. (1966). Hereditary variation of erythrocytic 6PGD. J. Lab. Clin. Med. **67**; 255.

Diamant S., Gobin E. and Shafir E. (1972). Enzyme activities related to fatty acid synthesis in liver and adipose tissue of rats treated with triiodothyronine. Eur. J. Biochem. **26**; 553.

Dickens F. (1936). Mechanism of carbohydrate oxidation. Nature **138**; 1057.

Donis-Keller H., Green P., Helms C., Cartinhour S., Weiffenbach B., Stephens K., Keith T.P., Bowden D.W., Smith D.R., Lander E.S., Botstein D., Akots G., Rediker K.S., Gravius T., Brown V.A., Rising M.B., Parker C., Powers J.A., Watt D.E.(1987). A genetic linkage map of the human genome. Cell **51**; 319.

Donis-Keller H., Helms C., Green P., Riethman H., Ramachandra S., Falls K., Bowden D.W., Weiffenbach B., Keith T., Stephens K., Canizzaro L.A., Shows T.B., Stewart G.D., Van Keuren M. (1989) A human genome linkage map with more than 500 RFLP loci and average marker spacing of 6cM. HGM10 Cytogenet. Cell Genet. **51**; 991.

Dracopoli N.C., Harnett P., Bale S.J., Stanger B.Z., Tucker M.A., Housman D.E. and Kefford R.F. (1989). Loss of alleles from the distal short arm of chromosome 1 occurs late in melanoma tumor progression. Proc. Natl. Acad. Sci.USA **86**; 4614.

Dracopoli N.C., Stanger B.Z., Ito C.Y., Call K.M., Lincoln S.E., Lander E.S. and Houseman D.E. (1988). A genetic linkage map from PND to FY on the short arm of human chromosome 1. Am. J. Hum. Gen. **43**; 462.

Drouin J. (1980). Cloning of human mitochondrial DNA in *E.Coli.* J.Mol.Biol. 5; 467.

Eaton J.W., Boraas M, and Etkin N.L.(1972). in "Hemoglobin and Red Cell Structure and Function" (G.Brewer ed.) p 121. Plenum Press, New York

Eaton J.W. and Brewer G.J. (1974) in "The Red Blood Cell" (Surgenor D.M., ed.) Academic Press, New York.

Egeland J.A., Gerhard D.S., Pauls D.L., Sussex J.N., Kidd K.K., Allen C.R., Hostetter A.M. and Housman D.E. (1987). Bipolar affective disorders linked to DNA markers on chromosome 11. Nature **325**; 783.

Eggleston L.V. and Krebs H.A. (1974). Regulation of the pentose phosphate cycle. Biochem. J. **138**; 425.

Evans H.J., Buckland R.A. and Pardue M.L. (1974). Location of the genes coding for 18S and 28S ribosomal RNA in the human genome. Chromosoma **48**; 405.

Evans R.M.(1988). The steroid and thyroid hormone receptor superfamily. Science **240**; 889-895.

Feinberg A.P. and Vogelstein B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Analyt. Biochem. **137**; 266.

Fildes and Parr (1963). Human red-cell 6-phosphoglucnate dehydrogenases. Nature **200**; 890.

Fong C.-T., Dracopoli N.C., White P.S., Merill P.T., Griffith R.C., Housman D.E. and Brodeur G.M. (1989). Loss of heterozygosity for the short arm of chromosome 1 in human neuroblastomas: Correlation with *N-myc* amplification. Proc. Natl. Acad. Sci. USA **86**; 3753.

Fothergill-Gilmore L.A. (1986). The evolution of the glycolytic pathway. Trend. Bio. **11**; 47.

Frischauf A.-M., Lehrach H., Poustka A. and Murray N. (1983). Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**; 827.

Garcia D.L. and Holten D. (1975). Inhibition of rat liver glucose-6-phosphate synthesis by glucagon. J. Biol. Chem. **250**; 3960.

Garcia G., Nogueira M. and Freire M. (1989). Purification and characterization of a cofactor that controls the oxidative phase of the pentose phosphate cycle in liver and other tissues of rat. Bioch. Biophys. Acta **990**; 59.

Gibson D.M., Lysons R.T., Scott D.F. and Muto Y. (1972). Synthesis and degradation of the lipogenic enzymes of rat liver. Adv. Enz. Reg. **10**; 187.

Giallongo A., Feo S., Moore R., Croce C.M. and Showe L.O. (1986). Molecular cloning and nucleotide sequence of a full-length cDNA for human α-enolase. Proc. Natl. Acad. Sci. **83**; 6741.

Glock G.E. and McLean P. (1954). Levels of enzymes of the direct oxidative pathway of carbohydrate matabolism in mammalian tissues and tumours. Biochem. J. **56**; 171.

Glock G.E. and McLean P. (1955). A preliminary investigation of the hormonal control of the hexose monophosphate oxidative pathway. Biochem. J. **61**; 390.

Goff S.P., Gilboa E., Witte O.N. and Baltimore D. (1980). Structure of the Abelson murine leukaemia virus gene and the homologous cellular gene: studies with cloned virus DNA. Cell **22**; 777.

Gordon H., Keraan M.M. and Vooijs M. (1967). Variant of 6-PGD within a community. Nature **214**; 466.

Green S. and Chambon P. (1988). Nuclear receptors enhance our understanding of transcriptional regulation. Trends Genet. **4**; 309.

Gusella J.F., Wexler N.S., Conneally P.M., Naylor S.L., Anderson M.A., Tanzi R.E., Watkins P.C., Ottina K., Wallace M.R., Sakaguchi A.Y., Young A.B., Shoulson I., Bonilla E. and Martin J.B. (1983). A polymorphic DNA marker genetically linked to Huntingdon's disease. Nature **306**; 234.

Hamerton J.L. and Cook P.J.L. (1974). Report on the genetic constitutions of chromosomes 1 and 2. Rotterdam Conference,1974: HGM2. Birth Defects Original Article Series XI,3;3.

Harper M.E. and Saunders G.F. Localisation iof single copy sequences on G-banded human chromosomes by *in situ* hybridisation. Chromosoma **83**; 431.

Harris H. and Hopkinson D.A. (1976). "Handbook of Enzyme Electrophoresis in Human Genetics" Amsterdam: North-Holland.

Hawthorne G.C. and Alberti K.G.M.M. (1988). The effect of high glucose and high insulin concentrations on pentose phosphate shunt enzymes and malic enzyme in cultured human endothelial cells. Horm. Metabol. Res. **20**; 645.

Harris H., Hopkinson D.A. and Robson E.B. (1962). Two-dimensional electrophoresis of pseudocholinesterase components in human serum. Nature **196**; 1296.

Henikoff S. and Smith J.M. (1989). The human mRNA that provides the n-terminus of chimeric G6PD encodes GMP reductase. Cell **58**; 1021.

Hodgkinson S., Sherrington R., Gurling H., Marchbanks R., Reeders S., Mallet J., McInnes M., Petersson H., Brynjolfsson J. (1987). Molecular genetic evidence for heterogeneity in manic depression. Nature **325**; 805.

Hopkinson D.A. (1970). The investigation of reactive sulphydryls in enzymes and their variants by starch gel electrophoresis: Studies on the human phosphohexose isomerase variant, PHI5-1. Ann. Hum. Genet. **34**; 79.

Hopkinson D.A. and Harris H. (1969). The investigation of reactive sulphydryls in enzymes and their variants by starch gel electrophoresis: Studies on red cell adenosine deaminase. Ann. Hum. Genet. **33**; 81.

Hopkinson D.A., Spencer N. and Harris H. (1963). Red cell acid phosphatase variants: A new human polymorphism. Nature **197**; 969.

Horecker B.L., Gibbs M., Klenow H. and Smyrniotis P.Z. (1954). The mechanism of pentose phosphate conversion to hexose monophosphate I. With liver enzyme preparation. J. Biol. Chem. **207**; 393.

Horecker B.L., Smyrniotis P.Z., Seegmiller J.E. (1951). Phosphogluconic acid dehydrogenase from yeast. J. Biol. Chem. **193**; 371.

Horecker B.L., Smyrniotis P.Z., Seegmiller J.E. (1951). The enzymatic conversion of 6-phosphogluconate to ribulose-5-P and ribose-5-P. J. Biol. Chem. **193**; 383.

Horecker B.L. and Mehler A.H. (1955). Carbohydrate metabolism. Ann Rev. Biochem. **24**; 207.

Huggins C. and Yao F. (1959). Influence of hormones on liver. I. Effects of steroids and thyroxine on pyridine nucleotide-linked dehydrogenases. J. Exp. Med. **110**; 899.

Hulten M. (1974). Chiasma distribution at diakinesis in the normal human male. Hereditas **76**; 55.

Hulten M., Palmer R.W. and Laurie D.A. (1982). Chiasma derived genetic maps and recombination fractions: Chromosome 1. Ann. Hum. Genet. **46**; 167.

Hutchison J. and Holten D, (1978). Quantitation of mRNA levels for rat liver 6-PGD. J.Biol. Chem. **253**; 52.

Jacob H.S. and Jandle J.H.(1966). Effect of sulfhydryl inhibition on red cells. J.Biol.Chem. **241**; 4243.

Jeffreys A.F. (1979). DNA sequence variants in the G -, A -, δ- and β-globin genes of man. Cell **18**; 1.

Jeffrey A.J., Wilson V. and Thein s.l.(1985). Hypervariable 'mimisatellite' regions in human DNA. Nature **314**; 67.

Jeffrey J., Soduling-Barros J., Murray L.A., Hansen R.J., Szepesi B., Jornvall H. (1988). Molecular diversity of glucose-6-phosphate dehydrogenase: Rat enzyme structure identifies NH_2-terminus seqment, shows initiation from nonequivalent sites in different organisms and establishes otherwise extensive conservation. Proc. Natl. Acad. Sci. USA 85; 7840.

Jenkins T. and Nurse G.T. (1974). The red cell 6-phosphogluconate dehydrogenase polymorphism in certain Southern African populations; with the first report of a new phenotype. Ann. Hum. Genet. **38**; 19.

Jones E.A., Goodfellow P.N., Kennett R.H. and Bodmer W.F. (1976). The independent expression of HLA and β-2-microglobulin in human-mouse hybrids. Somatic Cell Genet. **2**; 483.

Kan Y.W. and Dozy A.M. (1978). Polymorphism of DNA sequence adjacent to human β-globin structural gene: Relationship to sickle cell mutation. Proc. Natl. Acad. Sci. USA **75**; 5631.

Kanno H., Huang I-Y., Kan Y.W. and Yoshida A. (1989). Two structural genes on different chromosomes are required for encoding the major subunit of human red cell glucose-6-phosphat dehydrogenase. Cell **58**; 595.

Kather H., Rivera M. and Brand K. (1972). Interrelationship and control of glucose metabolism and lipogenesis in isolated fat cells. Biochem. J. **128**; 1097.

Katz J. and Rognstad R. (1966). The metabolism of tritiated glucose by rat adipose tissue. J. Biol. Chem. **241**; 3600.

Katz J. and Wals P.A. (1972). Pentose cycle and reducing equivalents in rat mammary gland slices. Biochem.J. 128; 879.

Keats B., Ott J. and Conneally M. (1989). Report of the committee on linkage and gene order. Cytogenet. Cell Genet. 51; 459.

Kelsoe J.R., Ginns E.I., Egeland J.A., Gerhard D.S., Goldstein A.M., Bale S.J., Pauls D.L., Long R.T., Kidd K.K., Conte G., Housman D.E. and Paul S.M. (1989). Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. Nature 342; 238.

Kidd K.K., Bowcock A>M>, Schmidtke J., Track R.K., Ricciuti F., Hutchings G., Bale A., Pearson P., Willard H.F. (1989). Report of the DNA commitee and catalogs of cloned and mapped genes and DNA polymorphisms. Cytogenet. Cell Genet. 51; 622.

Kielty C.M., Povey S. and Hopkinson D.A. (1982). Regulation of expression of liver specific enzymes III. Further analysis of a series of rat hepatoma and human somatic cell hybrids. Ann. Hum. Gen. 46; 307.

Kirkman H.N. and Gaetani G.F. (1986). Regulation of glucose-6-phosphate dehydrogenase in man. J. Biol. Chem. 261; 4033.

Kirkman H.N., Gaetani G.F., Clemons E.H. (1986) NADP-binding proteins causing rduced availability and sig-moid release of NADP in human erythrocytes. J. Biol. Chem. 261; 4039.

Kirkman H.N., Wilson W.G. and Clemons E.H. (1980). Regulation of glucose-6-phosphate dehydrogenase. J. Lab. Clin. Med. 95; 877.

Lalley, P.A., Davidson M.T., Graves J.A.M., O'Brien S.J., Womack J.E., Roderick T.H., Creau-Goldberg N., Hillyard A.L., Doolittle D.P. and Rogers J.A. (1989). Report of the committee on comparative mapping. Cytogenet. Cell Genet. 51; 503.

Landau B.R. (1989). Why the L-type pentose pathway does not function in liver. Int. J. Biochem. 21; 99.

Lander E.S. and Botstein D. (1986). Strategies for studying heterogeneous genetic traits in humans by using a linkage map of the restriction length polymorphisms. Proc. Natl. Acad. Sci. USA 83; 7353.

Lawn R.M., Fritsch E.F., Parker R.C., Blake G. and Maniatis T. (1978). The isolation and characterization of a linked δ+β globin gene from a cloned library of human DNA. Cell **15**; 1157.

Lewis W.H.P., Corney G. and Harris H. (1968). PepA5-1 and PepA6-1: Two new variants of peptidase A with featuraes of special interest. Ann. Hum. Genet. **32**;35 (1968).

Lippmann F. (1936). Fermentation of phosphogluconic acid. Nature **138**; 588.

Maniatis T., Fritsch E.F., Sambrook J. (1982). Molecular Cloning. A laboratory manual. Cold Spring Harbor Laboratory, New York.

Maniatis T., Goodbourn S., Fischer J.A. (1987). Regulation of inducible and tissue-specific gene expression. Science **236**; 1237.

McDonald, B.E., and Johnson B.C. Metabolic response to realimentation following chronic starvation in the adult male rat. J.Nutr. **87**; 161.

Mathew C.G.P., Smith B.A., Thorpe K., Wong Z., Royle N.J., Jeffreys A.J. and Ponder B.A.J. (1987). Deletion of genes on chromosome 1 in endocrine neoplasia. Nature **328**; 524.

Meera Khan P., Hagmeijer A., Wijnen L.M.M., van den Goes, R.G.M. (1984). PGM3 and ME1 are probably in the 6pter-q12 region. Cytogenet. Cell Genet. **37**; 537.

Messing J., Crea R. and Seeburg P.H. (1981). A system for shotgun DNA sequencing. Nucl. Acid Res. **9**; 309.

Miksicek R. and Towle H. (1982). Changes in the rates of synthesis and mRNA lavels of hepatic G6PD and PGD following induction by diet or thyroid hormone. J. Biol. Chem. **257**; 11829.

Miksicek R. and Towle H. (1983). Use of a cloned cDNA sequence to measure changes in PGD mRNA levels caused by thyroid hormone and dietary carbohydrate. J. Biol. Chem. **258**; 9575.

Mitchell P.J. and Tijan R. (1989). Transcriptional regulation in mammalian calls by sequence-specific DNA binding proteins. Science **245**; 371.

Moos M. and Gallwitz D. (1982). Structure of a human β-actin-related pseudogene which lacks intervening sequences. Proc. Natl. Acad. Sci. USA **78**; 1110.

Morton N.E. (1955). Sequential tests for the detection of linkage. Am. J. Hum. Genet. 7; 277.

Mourant A.E., Kopec A.C. and Domaniewska-Sobczak. (1976). "The distribution of the Human Blood Groups and Other Polymorphisms" 2nd edn. Oxford University Press. New York.

Muller H.J. (1916). The mechanism of crossing-over. Am. Nat. **50**; 193.

Nace C.S., Szepesi B. and Michaelis D.E. (1979). Regulation of glucose-6-phosphate dehydrogenase and malic enzyme in liver and adipose tissue: effect of dietary trilinolein level in starved-refed and *ad libitum*-fed rats. J. Nutr. **109**; 1094.

Nasoff M.S., Baker H. and Wolf R.E Jnr. (1984). Dna sequence of the *Escherichia Coli* gene, *gnd*, for 6-phosphogluconate dehydrogenase. Gene **27**; 253.

O'Connell P., Lathrop G.M., Nakamura Y., Leppert M.L., Ardinger R.H., Murray J.L., Lalouel J.M., White R. (1989). Twenty-eight loci form a continuous linkage map of markers for human chromosome 1. Genomics **4**; 12.

Oppenheimer J.H. and Samuels H.H.(eds). "Molecular Basis of Thyroid Hormone Action" Academic Press. New York (1983).

Pai G.S., Sprenkle J.A., Do T.T., Mareni C.E. and Migeon B.R. (1980). Localization of loci for hypoxanthine phosphoribosyltransferase and glucose-6-phosphate dehydrogenase and biochemical evidence of non-frandom X chromosome expression from studies of a human X-autosome translocation. Proc.Natl.Acad.Sci. USA **77**; 2810.

Parr C.W. (1966). Erythrocyte 6-phosphogluconate dehydrogenase polymorphism. Nature **210**; 487.

Parr C.W. and Fitch L.I. (1964). Hereditary partial deficiency of human erythrocyte 6-phosphogluconate dehydrogenase. Biochem. J. **93**; 28C.

Parr C.W. and Fitch L.I. (1967). Inherited quantitative variations in human phosphogluconate dehydrogenase. Ann. Hum. Genet. **30**; 339.

Parr C.W. and Parr I.B. (1965). Stability differences of inherited variants of human red cell PGD. Bichem. J. **95**; 16P.

Parrington J.M., West L.F. and Povey S.(1987). Loss of heterozygosity in hypotriploid cell cultures from testicular tumours. Hum. Genet. **77**; 269.

Petit C., Levilliers J. and Weissenbach J. (1988). Physical mapping of the human pseudo-autosomal region; comparison with genetic linkage map. EMBO J. 7; 2369.

Pinkel D., Straume T. and Gray J.W. (1986). Cytogenetic analysis using quantitative, high sensitivity, fluorescence hybridisation.

Procsal D., Winberry L and Holten D. (1976). Dietary regulation of 6-phosphogluconate dehydrogenase synthesis. J. Biol. Chem. **251**; 3539.

Raivio K.O. and Lazar C.S. (1981). The phosphogluconate pathway and synthesis of 5-phosphoribosyl-1-pyrophosphate in human fibrobasts. Bioch. Biophys. Acta **678**; 51.

Renwick J.H. (1971). The mapping of human chromosomes. Ann. Rev. Genet. **5**; 81.

Rognstad R. and Katz J. (1979). Effects of 2,4-dihdroxybutyrate on lipogenesis in rat hepatocytes. J.Biol. Chem. **254**; 11969.

Rudack D., Gozukara E.M., Chisolm E.M., Holten D. (1971). The effect of dietary carbohydrate and fat on the synthesis of rat liver 6-phosphogluconate dehydrogenase. Bioch. Biophys. Acta **252**; 305.

St.Clair D., Blackwood D., Muir W., Ballie D., Hubbarb A., Wright A. and Evans H.J. (1989). No linkage of chromosome 5q11-q13 markers to schizophrenia in Scottish families. Nature **339**; 305.

Sanger F., Nicklen S. and Coulson A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**; 5463.

Sap J., Muñoz A., Damm K., Goldberg Y.,Ghysdael J., Leutz A., Berg H. and Vennström B. The c-*erbA* protein is a high affinity receptor for thyroid hormone. Nature **324**; 635.

Sapag-Hagar M., Lagunas R. and Sols A. (1973). Apparent unbalance between the activities of 6-phosphogluconate and glucose-6-phosphate dehydrogenase in rat liver. Bioch. Biophys. Acta **50**; 179.

Schwartz D.C. and Cantor C.R. (1984). Separation of Yeast Chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell **37**; 67.

Sherrington R., Brynjolfsson J., Petrsson H., Potter M., Dudleston K., Barraclough B., Wasmuth J., Dobbs M. and Gurling H. (1988). Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature **336**; 164.

Shih L., Hsia D.Y., Bowman J.E., Shih S. and Shih P. (1968). The electrophoretic phenotypes of red cell 6-phosphogluconate dehydrogenase in Chinese populations. Am. J. Hum. Genet. **20**; 474.

Southern E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**; 503.

Sturtevant A.H. (1913). The linear arrangement of six sex-linked factors in *Drosophila,* as shown by their mode of association. J. Exp. Zool. **14**; 43.

Sun J.D. and Holten D. (1978) Levels of rat liver glucose-6-phosphate dehydrogenase messenger RNA. J.Biol. Chem. **253**; 6832.

Tashian R.E., Brewer G.J., Lehmann H., Davies D.A. and Rucknagel D.L. (1967). V. Further studies on Xavante Indians. Genetic variabilty in some serum and erythrocyte enzymes, hemoglobin and urinary excretion of β-aminoisobutyric acid. Am. J. Hum. Genet. **19**; 524.

Tepperman H.M and Tepperman J. (1963). On the response of hepatic glucose-6-phosphate dehydrogenase activity to changes in diet composition and food intake pattern. Adv. Enz. Reg. **1**; 121.

Tepperman H.M and Tepperman J. (1965). Effect of saturated fat diets on rat liver NADP-linked enzymes. Am. J. Physiol. **209**; 773.

Tills D., Van den Branden J.L., Clements V.R. and Mourant A.E. (1970). The distribution in man of genetic variants of 6-phosphogluconate dehydrogenase. Hum. Hered. **20**; 523.

Tills D., Van den Branden J.L., Clements V.R. and Mourant A.E. (1971). The distribution in man of genetic variants of 6-phosphogluconate dehydrogenase. Hum. Hered. **21**; 305.

Tomlinson J.E., Nakayama R. and Holten D. (1988). Repression of pentose phosphate pathway dehydrogenase synthesis and mRNA by dietary fat in rats. J.Nutr. **118**; 408.

Tuchinda S., Rucknagel D.L., Supa Na-Nakorn and Wasi P. (1968). The Thai variant and the distribution of alleles of 6-phosphogluconate dehydrogenase and the distribution of 6-phosphogluconate dehydrogenase deficiency in Thailand. Biochem. Genet. **2**; 253.

Twigg A.J. and Sherratt D. (1980). Trans-complementable copy-number mutants of plasmid ColE1. Nature **283**; 216-218.

Van Heyningen V., Bobrow M., Bodmer W.F., Gardiner S.E. and Hopkinson D.A. (1975). Chromosome assignment of some human enzyme loci: mitochondrial malate dehydrogenase to 7, mannosephosphate isomerase and pyruvate kinase to 15 and probably, esterase D to 13. Ann. Hum. Gen. **38**, 295.

Vanin E.F. (1985). Processed pseudogenes: Characteristics and evolution. Ann. Rev. Genet. **19**; 253.

Veech R.L., Eggleston L.V. and Krebs H.A. (1969). The redox state of free nicotinamide-adenine dinucleotid phophate in the cytoplasm of rat liver. Biochem. J. **115**; 609.

Wagner M, (1986). A consideration of the origin of processed pseudogenes. Trend. Genet. **2**; 134.

Wahl G.M., Stern M. and Stark G.R. (1979). Efficient transfer of large fragments from agarose gels to diazobenzyloxymethal-paper and rapid hybridisation by using dextran sulphate. Proc. Natl. Acad. Sci. USA **76**; 3683.

Wainwright B., Farall M., Watson E. and Williamson R. (1986). A model system for the analysis of gene exclusion : Cystic fibrosis and chromosome 19. J. Med. Genet. **23**; 417.

Warburg O., Christian W. and Griese A. (1935). WasseroffÜbertragendes Co-Ferment, seine Zusammensetzung und Wirkungsweise. Biochem Zeitung **282**; 157.

Weinberger C., Thompson C.C., Ong E.S., Lebo R., Gruol D.J. and Evans R.M. The *c-erbA* gene encodes a thyroid hormone receptor. Nature **324**; 641.

Weiss M.C. and Green H. (1967). Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. Proc. Natl. Acad. Sci. USA **58**; 1104.

Weitkamp l.R., Guttormsen S.A., Greendyke R.M. (1971). Genetic linkage between a locus for PGD and the Rhesus locus: Evaluation of possible heterogeneity in the recombination fraction between sexes and among families. Am. J. Hum. Gen. **23**; 462.

Williams J.F. (1980). A critical reexamination of the evidence for the reactions of the pentose pathway in animal tissues. Trends Biochem. **5**; 315.

Williams J.F., Arora K.K. and Longenecker J.P. (1987). Impediments which oppose acceptance of the classical (F-type) pentose cycle for liver, some neoplasms and photosynthetic tissue. The case for the L-type pentose pathway. Int. J. Biochem. **19**; 749.

Wood H.G. (1955). Significance of alternate pathways in the metabolism of glucose. Physiol. Rev. **35**; 841.

Wood H.G., Katz J., Landau B.R. (1963). Estimation of pathways of carbohydrate metabolism. Biochem Zeitung **338**; 809.