© 1990 Oxford University Press

The incidence and distribution of CpG→TpG transitions in the coagulation factor IX gene. A fresh look at CpG mutational hotspots

Peter M.Green, A.Jane Montandon, David R.Bentley, Rolf Ljung[1], Inga Marie Nilsson[1] and Francesco Giannelli

Paediatric Research Unit, Division of Medical and Molecular Genetics, 7th Floor Guy's Hospital Tower, London Bridge, London SE1 9RT, UK and [1]Department of Coagulation Disorders, University of Lund, General Hospital, S-214 01 Malmö, Sweden

Received March 3, 1990; Accepted May 14, 1990

ABSTRACT

The mutations of 76 haemophilia B patients representing the whole population registered with the Malmö haemophilia centre (42) and referrals from the UK, were characterised. RFLP haplotype analysis of the defective genes indicated that 51 single base pair substitutions were definitely of independent origin and 27 of these were CpG → TpG or CpA transitions. This represents a 38-fold excess over other single-base changes. Most of such transitions (82%) occur at 9 CpG sites occupying critical positions (transitions at 3 sites substitute essential arginines, while at 6 sites transition to TpG creates stop codons). Sixteen of the 18 possible transitions at these 9 sites cause clear haemophilia B and should be fully ascertained in our haemophilia B population. This allowed the direct estimate of the rate of CpG transitions. This is 1.05×10^{-7} substitutions per base per gamete per generation. The marked excess of CpG transitions in haemophilia B appears partly due to the high proportion of CpG sites at critical positions (at least 9 out of 20). We propose that this follows from the fact that male hemizygosity makes X-linked genes particularly susceptible to selective forces that tend to fix CpG sites arising at critical positions.

INTRODUCTION

Mutations that seriously impair the function of the gene for coagulation factor IX cause haemophilia B. This is an X-linked recessive disease maintained in the population by an equilibrium between the mutation rate of haemophilia B and the reduced genetic fitness of affected males. Unrelated patients with the disease therefore usually carry mutations of independent origin (1). The development of rapid procedures for the identification of all sequence variations in DNA has enabled us to undertake the characterisation of the haemophilia B mutations at a population level (2,3) and it has thus provided an excellent opportunity to re-examine the problem of the intrinsic mutability of specific residues such as methylatable cytosines.

Single base substitutions in genomic DNA may result from misreplication of normal DNA; from misrepair; or, finally, from replication of unrepaired templates damaged by spontaneous degradation or by exogenous physical and chemical agents (4).

The main types of degradative events are base loss and deamination of bases (4). The glycosylic bond between purines and the sugar phosphate backbone of the DNA is relatively labile and extrapolation from data on the depurination of naked double stranded DNA has suggested that a mammalian cell might lose 10,000 purines per day or cell generation (5). Depyrimidination is 20 times less frequent than depurination (6). Cytosine (C), adenine (A) and guanine (G) degrade to uracil (U), hypoxanthine (H) and xanthine (X) respectively through loss of their exocyclic amino group. However, A and G appear to deaminate at a rate four orders of magnitude lower than that of depurination (7). Purine deamination is therefore expected to contribute little to the instability of DNA. By contrast the deamination of C is a frequent event. In naked, single stranded DNA, deamination of C occurs at a rate similar to depurination, but the amino group of C is much more stable (> 100×) in double stranded DNA (8). 5-methyl cytosine (^{Me}C) is more labile and deaminates to thymine (T) three to four times as frequently as C to U both in deoxynucleotides and DNA solutions (5,9). Since T is a normal component of DNA, the accurate repair of lesions due to the deamination of ^{Me}C to T entails the detection of G:T mismatches and their correction to G:C rather than A:T pairs. By contrast U and H in DNA can be recognised and excised directly by specific glycosylases that start the base excision repair pathway. Similarly base-free sites are recognised by endonucleases of the excision repair pathway and apurinic sites may possibly even be repaired by the direct insertion of the appropriate purines (4).

Coulondre et al (10) noted a 10-fold increase of C → T transition at ^{Me}C in E. coli and attributed this to the relative difficulty of repairing the G:T mismatches. Recently, evidence has been presented to show that the repair of G:T mismatches is very biased in favour of restoring the G:C pair in both E. coli and mammalian cells (11,12). This, however, seems still inadequate to maintain ^{Me}C at the same level as other residues in the DNA. Thus methylatable CpG sites are rare in vertebrates, there is also a deficit of C + G (13,14), and the restriction sites containing CpG dinucleotides are frequently polymorphic (15).

Here we report the CpG → TpG transitions that we have

detected in the factor IX gene during our population studies, we consider the problem of the intrinsic mutability of ^{Me}C, and provide a direct estimate of the rate of CpG to TpG or CpA transitions.

MATERIALS AND METHODS

Blood samples were obtained from 76 haemophilia B patients: 42 from Sweden, 33 from the UK and one from Iceland. The Swedish patient designation and family numbers (in brackets) relate to one another as follows: Malmö 19 (14); Malmö 40 (180); Malmö 20 (651); Malmö 4 (281); Malmö 7 (661); Malmö 17 (64); Malmö 38 (633); Malmö 32 (229); Malmö 36 (267); Malmö 23 (636); Malmö 29 (228); Malmö 30 (646); Malmö 31 (177); Malmö 28 (309); Malmö 3 (110); Malmö 15 (611); Malmö 14 (241); Malmö 41 (282); Malmö 25 (139).

DNA was extracted by standard procedures (16). Then all the essential regions of the factor IX gene (promoter, exons, sequences important for the processing of the primary transcript) were isolated by PCR amplification of each patient's sample. Such regions were either directly sequenced as described (2) or screened for sequence changes by the amplification mismatch detection (AMD) method (3). These were then fully characterised by direct sequencing of the relevant amplified region (2).

Restriction fragment length polymorphisms (RFLPs) were examined by restriction digestion or size estimation of amplified segments comprising the five most informative polymorphisms of the factor IX gene (Taq I, Dde I, Mnl I, Bam HI, Hha I) (1). The primers and procedures used for such analysis are described elsewhere (Green et al., in preparation).

RESULTS

The factor IX mutations of 76 patients with haemophilia B were characterised. Forty-two of these patients represent an unbiased sample, as they are the whole population of haemophilia B patients registered with the Malmö haemophilia centre and approximately two-thirds of all Swedish registered cases. The remaining group of 34 patients is slightly biased in favour of severely and moderately affected patients because they were referred from different centres where a special effort for the collection of mild cases has not yet been made.

The mutation in 65 of these patients is a single base pair substitution. Of these mutations 36 are CpG → TpG or CpA transitions (presumably CpG → TpG transitions in the sense or antisense strand, respectively, of the factor IX gene). These mutations are listed in Table I. Since several mutants show identical base substitutions and might not represent independent mutations, RFLPs were examined in all genes showing identical mutations in order to identify those that could definitely be

Table 1. Mutations at CpG sites of factor IX characterized in this laboratory.

Mutation name	Exon	Change	Nuc Pos	Amino acid change	FIX C	FIX Ag
Malmö 6 (2)	b	C→T	6364 *	Arg-4→Trp *	<1 **	30 **
Malmö 19	b	"	"	"	2	36
Malmö 40	b	"	"	"	1	30
Malmö 20	b	"	"	"	2	27
UK 3 (2)	b	G→A	6365	Arg-4→Gln	1	48
UK 4 (2)	b	"	"	"	1	45
Malmö 4 § (2)	b	C→T	6460	Arg29→Stop	<1	<0 1
UK 14 (3)	b	"	"	"	<1	<2
UK 24	b	"	"	"	1	
UK 27	d	G→A	10430	Gly60→Ser	10	
Malmö 7 (40)	e	C→T	17761	Arg116→Stop	<1	<0 1
UK 23	f	C→T	20413	Arg145→Cys	1 8	43
UK 21	f	"	"	"	1	
Malmö 17	f	G→A	20414	Arg145→His	4–11	91
Malmö 38	f	"	"	"		
Malmö 32	f	"	"	"	6	115
Malmö 36	f	"	"	"	7	110
Malmö 23	f	"	"	"	7	148
Malmö 29	g	G→A	30150	Ala233→Thr	5–22	12
Malmö 30	g	"	"	"	12	
Malmö 31	g	"	"	"	11	15
Malmö 28	g	"	"	"	22	
UK 26	h	C→T	30863	Arg248→Stop	<1	4
Malmö 3 § (2)	h	"	"	"	<1	<0 1
Malmö 15	h	"	"	"	<1	<0 1
Malmö 14	h	"	"	"	<1	<0 1
Malmö 41	h	C→T	30875	Arg252→Stop	<1	6
Malmö 25	h	C→T	31008	Thr296→Met	4	15
UK 32	h	"	"	"	6	5
UK 34 ¶	h	C→T	31118	Arg333→Stop	6	
UK 2 (2)	h	G→A	31119	Arg333→Gln	<1	135
Iceland 1 (2)	h	"	"	"	3	119
UK 5 (2)	h	"	"	"	1–2	75
UK 18	h	"	"	"	1	130
UK 20	h	C→T	31133	Arg338→Stop	1.7	<1
UK 31	h	"	"	"	<1	

* Nucleotide and amino acid numbering as in ref 35. Factor IX activity (FIX:C) and antigen (FIX:Ag) measurements for the UK patients were supplied by the referring haemophilia centres. They are in iu/dl. § anti-factor IX antibodies found in patient (inhibitors) ¶ female haemophiliac heterozygous for this mutation

considered of independent origin, that is: new mutations or mutations in genes with different assortments of polymorphic markers. Fifty-one single base pair substitutions could definitely be considered of independent origin, and of these 27 were CpG → TpG or CpA transitions. Since there are only 20 CpG sites in the 1383 bp of the coding region of factor IX (ie 40 possible sites for C to T changes in total) the CpG → TpG or CpA transitions are in a 38-fold excess: that is, 27 in 40 versus 24 other substitutions in 1343.

The chance of detecting a single base pair substitution in the essential region of the factor IX gene, in a population of haemophilia B patients, is largely determined by the functional impairment it causes. We have therefore examined the likely consequences of the CpG → TpG or CpA transitions at each CpG site in the coding region of the factor IX gene and the frequency of mutants at such sites (Table II). There are nine CpG sites in the coding region of the factor IX gene that can, *a priori*, be thought to occupy critical positions. At three of these sites transition to TpG or CpA causes the loss of arginines at key sites for the maturation and activation of factor IX: Arg −4, +145 and +180, and transition to TpG in the remaining 6 sites creates stop codons (codon 29, 116, 248, 252, 333, 338). There is, in fact, factual evidence that Arg −4, +145 and +180 are essential (Table I and refs.17−19). Furthermore, four CpG → CpA transitions in the non-coding strand of the sites that can form stop codons upon transition to TpG (29, 248, 333 and 338) cause Arg → Gln substitutions that have been found in patients with moderate or severe haemophilia B and thus shown to have serious detrimental consequences (20−23). In the population we have analysed we find 6 independent transitions (4 TpG and 2 CpA) at the CpG site in codon −4, and 4 such transitions (2 of each type) at the CpG site in codon +145. Furthermore we find ten independent transitions causing stop codons (ie: 3, 1, 3, 1, 1, 1 at codon 29, 116, 248, 252, 333 and 338 respectively) and 3 causing Arg → Gln substitution at codon 333.

A priori, the importance of the position of the other 11 CpG sites is less clear. Consideration of the Dayhoff's matrix (24) suggests that two CpG → TpG transitions would cause non-

Table 2. All the possible CpG → TpG transitions in the coding region of factor IX

Exon	Nuc Position	Change*	Amino acid Change	Dayhoff No	Mutations Obs§	Mutations Ind¶	Base in other spp D	S	P	Rb	G	R	M **
a	36	C→T	Arg-44→Cys	−4			–	–	–	–	–	–	–
a	37	G→A	Arg-44→His	+2			–	–	–	–	–	–	–
a	38	C→T	Arg-44silent				–	–	–	–	–	–	–
a	39	G→A	Val-43→Met	+2			–	–	–	–	–	–	–
b	6345	C→T	Asn-11silent				T	–	–	–	–	–	–
b	6346	G→A	Ala-10→Thr	+1				–	–	–	–	–	–
b	6364	C→T	Arg-4→Trp	+2	4	4		–	–	–	–	–	–
b	6365	G→A	Arg-4→Gln	+1	2	2		–	–	–	–	–	–
b	6460	C→T	Arg29→STOP		3	3		–	–	–	–	–	–
b	6461	G→A	Arg29→Gln	+1				–	–	–	–	–	–
d	10429	C→T	Gly59silent					–	–	–	–	–	–
d	10430	G→A	Gly60→Ser	+1	1	1		–	–	–	–	–	–
e	17700	C→T	Cys95silent				T	–	–	–	–	–	–
e	17701	G→A	Glu96→Lys	0				–	–	–	–	–	–
e	17761	C→T	Arg116→STOP		1	1		–	–	–	–	–	–
e	17762	G→A	Arg116→Gln	+1			A	–	–	–	–	–	–
f	20413	C→T	Arg145→Cys	−4	2	2							A
f	20414	G→A	Arg145→His	+2	5	2							
f	20518	C→T	Arg180→Trp	+2									
f	20519	G→A	Arg180→Gln	+1									
g	30083	C→T	Ile210silent										
g	30084	G→A	Val211→Ile	+4			A		A	A	A	A	A
g	30149	C→T	Val232silent				T	T	T	T	T	T	T
g	30150	G→A	Ala233→Thr	+1	4	2							
h	30863	C→T	Arg248→STOP		4	3						A	A
h	30864	G→A	Arg248→Gln	+1									
h	30875	C→T	Arg252→STOP										
h	30876	G→A	Arg252→Gln	+1	1	1				A			
h	30949	C→T	Asp276silent				T	T	T	T		T	T
h	30950	G→A	Glu277→Lys	0						A	A	A	A
h	30973	C→T	Tyr284silent				T	T		T		T	T
h	30974	G→A	Val285→Ile	+4									
h	31008	C→T	Thr296→Met	−1	2	2							
h	31009	G→A	Thr296silent						C	A	A	A	A
h	31118	C→T	Arg333→STOP		1	1							A
h	31119	G→A	Arg333→Gln	+1	4	3							
h	31133	C→T	Arg338→STOP		2	1						A	A
h	31134	G→A	Arg338→Gln	+1									
h	31328	C→T	Arg403→Trp	+2							T	A	
h	31329	G→A	Arg403→Gln	+1									

* C → T, G → A = transitions in sense and antisense strand respectively
§ see table 1 for more details on observed mutations
¶ ie in genes of different haplotype (independent)
D = dog, S = sheep, P = pig, Rb = rabbit, G = Guinea pig, R = rat, M = mouse. ** (−) indicates sequence not known.
Only differences to human CpGs are presented (32,33)

conservative amino acid substitutions $Arg_{-44} \rightarrow Cys$ (value −4) and $Thr_{296} \rightarrow Met$ (value −1), but the position of the translation start is not absolutely defined and could be 3′ of codon −44. Two examples of the $Thr_{296} \rightarrow Met$ were found in our study. Two further transitions have a score of 0 in Dayhoff's table, that is: Glu → Lys at position 96 and 277. Such mutations are not represented in our sample. Nine transitions lead to substitutions with Dayhoff's positive scores (+2, +2, +1, +1, +4, +1, +4, +2, +1); that is: $Arg_{-44} \rightarrow His$; $Val_{-43} \rightarrow Met$; $Ala_{-10} \rightarrow Thr$; $Gly_{60} \rightarrow Ser$; $Val_{211} \rightarrow Ile$; $Ala_{233} \rightarrow Thr$; $Val_{285} \rightarrow Ile$; $Arg_{403} \rightarrow Trp$; $Arg_{403} \rightarrow Gln$. We have found only one example of the mutation at codon 60 and 2 independent examples of that at codon 233. The remaining 9 transitions are silent mutations and none occurred in our sample. Thus when the distribution of haemophilia B mutations due to transitions at CpG sites is considered, 82% are at the 9 sites that clearly occupy critical positions in the factor IX gene. By contrast the remaining 11 sites contribute only 18% of such transitions.

Mutations at the 9 CpG sites which occupy critical positions provide a direct estimate of the CpG → TpG or CpA transition rate because: all 6 stop codons, the substitutions of Arg −4 and 180 by Gln and Trp or Arg_{145} by His or Cys, and at least the substitutions of Arg by Gln at positions 29, 248, 333 and 338 are known to cause haemophilia (Table I and refs. 19−23,25). Therefore we can expect that our study of haemophilia B populations will have detected most of the CpG → TpG transitions at the 9 selected CpG sites. We have found 23 independent transitions at these nine sites in a group of 76 patients from a representative population of haemophilia B patients. Since the frequency of haemophilia B is 1/30,000 this represents a total of 2,280,000 factor IX genes. Furthermore since there are at least 16 CpG → TpG or CpA transitions at the 9 CpG sites that can certainly cause haemophilia B we have observed one such transition per 793,043 sites. In order to transform this figure into a transition rate per base per generation we have to consider that in an X-linked recessive condition with a selection coefficient of 0.5, as reported for haemophilia B, prior to the introduction of modern therapy (26), the incidence of the disease in males is equal to 6 times the mutation rate. This leads to an estimate of 1:9,516,521 or 1.05×10^{-7} CpG → TpG or CpA transitions per base, per gamete, per generation.

DISCUSSION

The results presented above show that the CpG sites of the factor IX gene do not contribute in equal measure to the excess of CpG → TpG transitions causing haemophilia B and confirm the expectation that the spectrum of mutations in haemophilia B will be strongly biased in favour of those that cause sufficiently severe functional impairment. Therefore the great excess of transitions at CpG sites in the factor IX gene of haemophilia B patients (20, this work) does not simply reflect the instability of ^{Me}C but also the frequency of CpG at critical positions.

Comparison of the factor IX gene with those for factor VII, X and protein C that are thought to derive from the same ancestral gene as factor IX shows not only that the factor IX gene has a high proportion of CpG sites at critical positions but also that these do not reflect an ancestral situation because factor VII, X and protein C have only 2, 0 and 2 sites, respectively, that can form stop codons by mutating to TpG, even though they are at least 3 times richer than factor IX in CpG sites (27−30). The high proportion of CpG sites at critical positions in the factor IX gene may, instead, be the result of selection pressure. The relative instability of ^{Me}C and hence of ^{Me}CpG implies that during evolution the position of CpG sites may be particularly prone to change as some sites are lost and others are gained by any given stretch of sequence. Functional constraints to such 'movement' can be expected to occur. Thus, for example, when a CpG → TpG transition generates a stop codon and abrogates gene function, natural selection will tend to prevent the CpG from evolving through a path initiating with the transition of C to T. The strength of such a constraint is determined by the probability that the detrimental mutation will persist in the population and eventually, through further mutations, become part of an acceptable functional sequence. Recessive detrimental mutations of X-linked genes are particularly exposed to natural selection because of the male's hemizygosity. Thus mutations in the factor IX gene are subjected to strong negative selection because the gene is X-linked in Man and other mammals where factor IX deficiency also causes haemophilia (31). Thus CpGs that were at or have come to occupy critical sites in the factor IX gene may be relatively 'fixed'

In keeping with this idea the complete DNA sequence of canine factor IX (32) and the sequences of the activation peptide and catalytic region (coding for amino acids 137 to 407) of the factor IX of sheep, pig, rabbit, guinea pig, rat and mouse (33) show that the 9 critical CpG sites of human factor IX are highly conserved (Table II). Only two CpG to TpG transitions are observed at these sites: one in the codon for Arg 116 and one in that for Arg 252 that are replaced by Gln in the dog and guinea pig respectively. Interestingly, transversions to ApG, that do not alter the amino acid sequence, are observed more frequently. That is, in the CpGs at codons 145 and 333 in the mouse, and 248 and 338 in the mouse and rat. By contrast the other eleven CpG sites of human factor IX are poorly conserved and mostly represented by sequences that could be derived by CpG transitions (see Table II).

Also in keeping with the above hypothesis is the 4-fold excess of CpG sites in the coding region of factor IX relative to the whole gene (2.9% versus 0.7%) and the 4.3-fold excess of Taq I sites in the coding region of the factor VIII gene that yield a stop codon by CpG → TpG transition (5 out of 7 versus 1 out of 12 expected by chance alone). The evidence for the enrichment of CpG sites at critical positions indicates that crude ratios of the frequency of CpG to TpG transitions to other single base pair substitutions causing haemophilia B and other diseases (20, 36) may grossly overestimate the relative instability of CpG sites. In fact Sved and Bird (37) calculate that a rate of CpG → TpG transition equal to about 12 times the normal transition rate could account for an equilibrium where the frequency of CpG remains at the current level (20% of expected frequency).

The identification of 9 CpG sites at critical positions in the factor IX gene and the study of a representative population oof haemophilia B patients has ?? as directly to determine the rate of CpG transitions. If we convert this rate (1.05×10^{-7}) into a rate per base per year a value is obtained (5×10^{-9}) that is only 3−4 fold greater than the rate of sequence evolution of the DNA sequences of primates not subjected to conservational constraints: 1.4×10^{-9} (38,39).

ACKNOWLEDGEMENTS

We thank the patients who donated blood for this study and we are grateful to our colleagues and collaborators: Dr. C. Berry,

Dr. E. Goldman, Dr. P.B.A. Kernoff, Dr. M. Makris, Dr. R.S. Mibashan, Prof. F.E. Preston, Dr. C.R. Rizza, Dr. D.N. Whitmore for their generous assisatance in this study. We are also indebted to the Medical Research Council, Action Research for the Crippled Child, the Generation Trust, the Spastics Society and the Swedish Medical Research Council for financial assistance.

REFERENCES

1 Giannelli F (1989) In Tuddenham, E G D (ed), Balliere's clinical haematology, Balliere, Tindall, London. **2** 821–848
2 Green, P M , Bentley, D R , Mibashan, R S., Nilsson, I.M and Giannelli, F (1989) EMBO J. **8** 1067–1072
3 Montandon, A.J , Green, P M , Giannelli, F and Bentley, D R.(1989) Nucl Acids Res. **17** 3347–3358.
4 Friedberg, E C. (1985) In DNA repair, W H Freeman & Co., New York
5 Lindahl, T. (1979) In Cohn, W.E. (ed), Academic Press, London, New York Prog. Nucleic Acids Res Mol. Biol **22**· 135–192
6 Lindahl, Y , & Karlstrom, O (1979) Biochemistry **12** 5151–5154.
7 Jordon, D O (1960) In The chemistry of nucleic acids, Butterworth, Washington DC
8 Lindahl, T & Nyberg, B. (1974) Biochemistry **13**: 3405–3410
9 Wang, R Y-H, Kuo, K C , Gehrke, C W , Huang, L-H & Ehrlich, H. (1982) Biochim Biophys. Acta. **697**: 371–377.
10 Coulondre, C , Miller, J H., Farabauch, P J & Gilbert, W (1978) Nature **274** 775–780
11 Zell, R & Fritz H-J (1987) EMBO J **6** 1809–1815.
12 Brown, T C & Jincny, J. (1988) Cell **54** 705–711
13 Setlaw, P (1976) In Fasman G D (ed) CRC handbook of biochemistry and molecular biology, vol 2 nucleic acids 3rd edition, CRC press
14 Bird, A (1980) Nucleic Acids Res **8**· 1499–1504
15 Barker, D , Schafer, M and White, R (1984) Cell **36** 131–138
16 Maniatis, T , Fritsch, E F , & Sambrook, J (1982) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
17 Bentley, A K , Rees, D.J G , Rizza, C , and Brownlee, G.G (1986) Cell **45** 343–348.
18 Noyes, C M , Griffith, M J , Roberts, H.R and Lundblad, R L (1983) Proc Natl. Acad Sci , USA **80** 4200–4202.
19 Huang, M.N., Kasper, C K , Roberts, H R , Stafford, D W & High, K.A (1989) Blood **73** 718–721
20. Koeberl, D D., Bottema, C D K , Buerstedde, J.M & Sommer, S.S (1989) Am J Hum Genet **45**. 448–457
21 Chen, S H., Thompson, A R , Zhang, M & Scott, C R (1989) J Clin Invest. **84**. 113–118.
22 Tsang, T C., Bentley, D R , Mibashan, R.S & Giannelli, F (1988) EMBO J **7** 3009–3015.
23 Thompson, A R (1990) Prog. Haemostas Thrombos. 10: (in press)
24 Dayhoff, M O , Schwartz, R M and Orcut, B C (1979) In Atlas of protein sequence and structure, vol 5, suppl. 3, National Biomedical Research Foundation, Washington.
25. Suehiro, K , Kawabata, S , Miyata, T , Takeya, H , Takamatsu, J , Ogata, K , Kamiya, T , Saito, H., Niho, Y. & Iwanaga, S. (1989) J Biol Chem. **264**: 21257–21265
26 Ferrari, N and Rizza, C A. (1986) Brazilian J. Genet **9** 87–99.
27 O'Hara, P J., Grant, F.J , Haldman, B.A , Gray, C.L., Insley, M Y., Hagen, F S & Murray, M.J (1987) Proc. Natl. Acad. Sci , USA **84** 5158–5162.
28 Leytus, S.P , Chung, D W., Kisiel, W., Kurachi, K. & Davie, E.W. (1984) Proc. Natl Acad. Sci , USA **81** 3699–3702.
29. Foster, D.C., Yoshitake, S. & Davie, E W. (1985) Proc. Natl Acad Sci., USA **82** 4673–4677.
30 Davie, E W. (1986) Chem Scr **26B**: 241–245
31 Evans, J.P., Brinkhous, K.M , Brayer, G D., Reisner, H.M & High, K A. (1989) Proc. Natl Acad Sci , USA **86** 10095–10099.
32 Evans, J.P., Watzke, H.H., Ware, J.L , Stafford, D W & High K.A (1989) Blood **74**: 207–212.
33. Sarkar, G., Koeberl, D D & Sommer, S.S (1990) Genomics **6** 133–143
34. Anson, D S , , Choo, K.H , Rees, D.J.G., Giannelli, F , Gould, K , Huddleston, J.A., Brownlee, G.G. (1984) EMBO J. **3**. 1053–1053
35 Yoshitake, S., Schach, B.G., Foster, D.C , Davie, E.W., and Kurachi, K (1985) Biochemistry **24**: 3736–3750.
36 Cooper, D.N. & Krawczak, M (1989) Hum Genet. **83** 181–188
37. Sved, J & Bird, A (1990) Proc Natl. Acad Sci , USA (in press)
38 Koop, B F , Goodman, M., Xu, P., Chan, K. & Slightom, J.L. (1986) Nature **319** 234–238.
39 Miyamoto, M.M , Koop, B.F., Slightom, G.L., Goodman, M. & Tennant, M R. (1988) Proc Natl Acad. Sci , USA **85**: 7627–7631
40 Montandon, A J , Green, P M., Bentley, D R , Ljung, R., Nilsson, I.M. & Giannelli, F B (1990) Hum. Genet (in press)