Evolutionary and structural relationships among the group-specific component, albumin and α -fetoprotein

Funmei Yang, Victor J.Luna, Robert D.McAnelly, Kathleen H.Naberhaus, Rod L.Cupples and Barbara H.Bowman

Department of Cellular and Structural Biology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78284, USA

Received 28 August 1985; Accepted 21 October 1985

ABSTRACT

The group-specific component (Gc) is a plasma protein that binds vitamin D. Recent characterization of human *Gc* cDNA demonstrated homology with serum albumin and α -fetoprotein. This study compares the sequences of the three proteins and demonstrates a strong evolutionary relationship. Albumin, α -fetoprotein and *Gc* evolved from an ancestral gene containing an intragenic triplication. Comparison of the amino acid sequences and patterns of double disulfide bonds suggests that the *Gc* gene may have diverged from an ancestral gene earlier in evolution than the genes encoding albumin and α -fetoprotein. Analysis of the amino acid and nucleotide sequences of the three internal domains of *Gc* revealed 19-23% amino acid sequence identity and the localization of three homology blocks with 40-44% nucleotide sequence identity. The deduced amino sequence of *Gc* furnished data for comparing its molecular configuration based on the predicted secondary structure with those predicted for human albumin and α -fetoprotein. Utilization of *Gc* cDNA has also led to the identification of its genomic DNA and detection of a human DNA polymorphism.

INTRODUCTION

The group-specific component (Gc) is the major vitamin D binding protein in plasma (1). It is of genetic interest because three common genetic phenotypes, Gc1, Gc2 and Gc 2-1, appear in all human populations (2). Gc1 and Gc2 differ by at least three amino acid residues (3), including a Thr residue in Gc1 that is glycosylated and carries a sialic acid group that is absent in the homologous region of Gc2 (3,4,5).

The *Gc* gene is located approximately 1.5 centimorgans from the genes encoding albumin and α -fetoprotein on the long arm of human chromosome 4 (6,7,8). We recently characterized human *Gc2* cDNA and demonstrated that *Gc* shares sequence homology with human albumin and α -fetoprotein (9). All three plasma proteins appear to share a common ancestor that was the product of intragenic triplication.

MATERIALS AND METHODS

Materials

Restriction endonucleases, polymerases, radiolabeled compounds and other reagents utilized in these experiments have been specified previously (10).

Analysis of Sequence Homology

Homologies in α -fetoprotein, albumin and Gc were observed after aligning their amino acid sequences according to the positions of homologous double disulfide bonds (11) and introducing gaps to maximize homology. Computer analysis of the amino acid sequences was performed using the programs ALIGN and SEARCH provided by the National Biomedical Research Foundation. The protein sequence database of the Atlas of Protein Sequence and Structure (updated to May 1985) was searched with a 27 residue segment of Gc to identify all possible related sequences. The ALIGN program was then used to determine the best possible alignment of related proteins by scoring amino acid pairing according to the mutation data matrix. Within the program, gaps were inserted to maximize homology but limited by use of a gap penalty of 8. The statistical significance of each best alignment was determined by computing the distribution of scores for 100 random alignments and calculating an alignment score. The alignment score was the number of standard deviations (SD) by which the maximum score exceeded the mean score of the chance permutations. Alignment of the three internal domains of Gc was performed in a similar manner.

Homology of nucleotide sequences in the three domains of Gc was evaluated by calculation of "accident probability," P_a , the probability that a homology equal to, or greater than, that being considered might arise accidentally (12). A homology observed, within a block comparing two sequences of over 100 nucleotides each, with a P_a of $<7 \times 10^{-4}$ was considered significant.

Prediction of Secondary Structure from the Deduced Amino Acid Sequence

The procedures of Chou and Fasman (13,14) were used to predict the secondary structure of human Gc, albumin, and α -fetoprotein.

Preparation of Probes

The probe used to identify the Gc genomic DNA clones was prepared from clone Gc2p isolated by Yang et al. (9). Clone Gc2p contains the cDNA encoding the entire Gc2 sequence. Plasmid DNA from clone Gc2p was digested with *Pst*I restriction endonuclease and separated by agarose gel electrophoresis. The 1.8 kbp insert fragment was isolated by electroelution. The fragment was then labeled with $\alpha^{32}\text{p}$ dCTP by using oligonucleotides as primers following the procedure of Feinberg and Vogelstein (15).

Southern Blot Analysis of DNA Polymorphisms

High molecular weight DNAs were isolated from human lymphocyte nuclei by the method of Bell et al. (16). Filter hybridization of lymphocyte DNA with human *Gc* cDNA was carried out at 42°C in 50% formamide, 5X SSPE, 10% dextran sulfate, 5X Denhart's solution and 0.1% SDS. Filters were washed with 2X SSC at room temperature followed by washing with 0.1X SSC plus 0.1% SDS at 55°C.

RESULTS AND DISCUSSION

The deduced sequence of human *Gc* had demonstrated strong homology with α -fetoprotein and albumin (9), especially in the distribution of double disulfide bonds characteristic of these proteins. The role of the double disulfide bonds in forming the loops in each of the homologous three domains of albumin was first demonstrated by Brown (11).

The homology existing in the amino acid sequences of *Gc*, albumin and α -fetoprotein is presented in Figure 1. After the amino acid sequences are aligned according to the double disulfide bonds, 11 gaps in *Gc*'s sequence and 6 gaps in both the α -fetoprotein and albumin amino acid sequences were introduced to maximize homology. The amino acid sequences of *Gc* and albumin were 23% identical; *Gc* and α -fetoprotein were 19% identical. In this alignment, albumin and α -fetoprotein were 35% identical.

Further confirmation of *Gc*'s placement in the albumin family was gained using the computer program SEARCH. When a 27 residue segment of *Gc*, residues 80-106, chosen because of containment of the unique double disulfide bonding, was used to search the protein database, albumin and α -fetoprotein were identified as significant matches from among over 3000 proteins searched.

Using the program ALIGN, the amino acid sequence of *Gc*, including the leader sequence, was aligned to portions of α -fetoprotein and albumin of comparable length. Residues -16 to 458 of *Gc* were aligned with residues -19 to 466 of α -fetoprotein to give an alignment score of 19.59 SD units and aligned with residues -18 to 461 of albumin for a score of 24.87 SD units. In general, an alignment score of >3 SD units may be considered statistically significant (17). Thus, the *Gc*, α -fetoprotein and albumin alignments not only exceed the requirement for significance but also meet the definition proposed by Doolittle (18) for grouping of protein families.

The leader and N-terminal sequences of *Gc*, albumin and α -fetoprotein were found to be highly similar irrespective of where the N-terminal sites occurred in the three homologous proteins. Figure 2 compares the leader and N-terminal regions of the three proteins. There is 40% identity in amino acid sequences

AFP	⁺¹ T L H R N E Y G I A S I L D S Y Q C T A E I S L A D L A T I F F A Q F V Q E A T Y K E V S K	46
Gc	⁺¹ L E R G R D Y E K N K V C K E F S H L G K E D F T S L S L V L Y S R K F P S G T F E Q V S Q	46
ALB	D A H K S E V A H R F K D L G E E N F K A L V L I A F A Q Y L Q Q C P F E D H V K	41
AFP	M V K D A L T A I E K P T G D E Q S S G C L E N Q L P A F L E E L C H E K E I L E - K Y G H	91
Gc	L V K E V V S L T E A C C A E G A D P D C Y D T R T S A L S A K S C E S N S P F P V H P G T	92
ALB	L V N E V T E F A K T C V A D E S A E N C D K S L H T L F G D K L C T V A T L R E T Y G E M	87
AFP	S D C C S Q S E E G R H N C F L A H K K P T P A S I P L F Q V P E P V T S C E A Y E E D R E	137
Gc	A E C C T K E G L E R K L C M A A L K H Q - P Q E F P T Y V E P T N D E I C E A F R K D P K	137
ALB	A D C C A K Q E P G R N E C F L Q H K D D N P N - L P R L V R P E V D V M C T A F H D N E E	132
AFP	T F M N K F I Y E I A R R H P F L Y A P T I L L W A A R Y D K I I P S C C K A E N A V E C F	183
Gc	E Y A N Q F M W E Y S T N Y G Q A P L S L L V S Y T K S Y L S M V G S C C T S A S P T V C F	183
ALB	T F L K K Y L Y E I A R R H P Y F Y A P E L L F F A K R Y K A A F T E C C Q A A D K A A C L	178
AFP	Q T K A A T V T K E L R E S S L L N Q H A C A V M K N F G T R T F Q A I T V T K L S Q K F T	229
Gc	L - K E R L Q L K H L S L L T T L S N R V C S Q Y A A Y G E K K S R L S N L I K L A Q K V P	228
ALB	L P K L D E L R D E G K A S S A K Q R L K C A S L Q K F G E R A F K A W A V A R L S Q R F P	224
AFP	K V N F T E I Q K L V L D V A H V H E H C C R G D V L D C L Q - D G E K I M S Y I C S Q Q D	274
Gc	T A D L E D V L P L A E D I T N I L S K C C E S A S E D G M A K E L P E H T V K L C D N L S	274
ALB	K A E F A E V S K L V T D L T K V H T E C C H G D L L E C - A D D R A D L A K Y I C E N Q D	269
AFP	T L S N K I T E C C K L - T T L E R G Q C I I H A E N D E K P E G L S P N L - N R F L G D R	318
Gc	T K N S K F E D C C Q E K T A M D V F V C T Y F M P A A Q L P E - L - P D V - - E L P T N K	316
ALB	S I S S K L K E C C - E K P L L E K S H C I A E V E N D E M P A D L - P S L A A D F V E S K	313
AFP	D F N Q F S S G E K N I F L A S F V H E Y S R R H P Q L A V S V I L R V A K G Y Q E L L E K	364
Gc	D V C D - - P G N T K V M - D K Y T F E L S R R T H L P E V F L S K V L E P T - L K S L G E	358
ALB	D V C K N Y A E A K D V F L G M F L Y E Y A R R H P D Y S V V L L L R L A K T Y E T T L E K	359
AFP	C F Q T E N P L E C Q D - K G E E L Q K Y I - Q E S Q A L A K R S C G L F Q K L G E Y Y L	408
Gc	C C D V E D S T T C F N A K G P L - L K K E L - S S F I D K G Q E L C A D Y S E N T F T E Y	402
ALB	C C A A A D P H E C Y - A K V F D E - F K P L V E E P Q N L I K Q N C E L F E Q L G E Y K F	403
AFP	Q N A F L V A Y T K K A P Q L T S S E L M A I T R K M A A T A A T C C Q L S E D K L L A C G	454
Gc	K K K L A E R L K A K L P D A T P K E L A K L V N K R S D F A S N C C S I N S P P L - Y C D	447
ALB	Q N A L L V R Y T K K V P Q V S T P T L V E V S R N L G K V G S K C C K H P E A K R M P C A	449
AFP	E G A A D I I G H L . . .	465
Gc	S E I D A E L K N I L	458
ALB	E D Y L S V V L N Q L . . .	460

Fig. 1. Homology of amino acids in sequences of human α -fetoprotein, Gc and albumin. Residues identical to Gc are boxed. Numbers to the right of the figure indicate amino acid residue positions.

AFP	-19	Met Lys Trp Val	Glu Ser	Ile Phe	Leu Ile	Phe Leu	Leu Asn	Phe Thr	Glu	Ser Arg	Thr	+1
		ATG AAG TGG GTC	GAA TCA	ATT TTT	TTA ATT	TTC CTA	CTA AAT TTT	ACT GAA	TCC AGA	ACA		
ALB	-24	Met Lys Trp Val	Thr Phe	Ile Ser	Leu Leu	Phe Leu Phe	Ser Ser	Ala Tyr	Ser Arg	Gly	-5	
		ATG AAG TGG GTA	ACC TTT	ATT TCC	CTT CTT	TTT CTC TTT	AGC TCG	GCT TAT	TCC AGG	GGT		
Gc	-16	Met Lys	Arg Val	Leu Val	Leu Leu	Leu Ala	Val Ala	Phe	Gly His	Ala	Leu Glu	Arg Gly
		ATG AAG	AGG GTC	CTG GTA	CTA CTG	CTT GCT	GTG GCA	TTT	GGA CAT	GCT	TTA GAG	AGA GGC

Fig. 2. Protein and nucleotide leader sequences of α -fetoprotein, albumin and Gc. Identical amino acid residues are boxed.

of Gc and albumin and 48% identity in respect to the nucleotide sequence in this region. When the same regions are compared in Gc and α -fetoprotein there was 25% identity found in the amino acid sequence and 42% identity found in the nucleotide sequence. In this region, albumin and α -fetoprotein demonstrate 50% identity in respect to amino acid sequence and 57% identity in nucleotide sequence.

A comparison of the amino acid sequences of the three domains of Gc confirms that Gc, like albumin and α -fetoprotein, is the evolutionary product of intragenic triplication, with 23% identity between domains I and II and 19% identity between II and III (Fig. 3). The Gc cDNA sequence demonstrated that

Domain I	L E R G R D Y E K N K V C K E F S H L G K E D F T S L S L V L Y S R K F P S G T F E Q V S	45
Domain II	L S L L T T L S N R V C S Q Y A A Y G E K K S R L S N L I K L A Q K V P T A D L E D V L	236
Domain III	F L S S F I D K G Q E L C A D Y S E N T F T E Y K K K I A E R L K A K L P D A T P K E L A	423
Domain I	Q I L V K E V V S L T F A C C A E G A D P D - C Y D T R T S A L S A K S C E S N S P F P V H	89
Domain II	P L A E D I T N I L S K C C - E S A S E D - C H A K E L P E H T V K L C D N I S T K N S -	288
Domain III	K I L V N K R S D F A S N C C S I N S P L Y C D S E I D A E L K N I L	458
Domain I	P G T A E C C - T K E G L E R K L C H A A L K - H Q P Q E F P T Y V E P T N D E I C E A F	132
Domain II	- K F E D C C Q E K T A M D V F V C T Y F M P A A Q L P E L P D V E L P T N K D V C D P G	322
Domain I	R K D P K E Y A N Q F M W E Y S T N Y G Q A P L S L L V S Y T K S Y L S H V G S C C T S A	188
Domain II	N T K V M D K Y T - F E L S R R T H L F E V F L S K - V L - E P T L K S - L G E C C D V E	363
Domain I	S P T V C F L K E R L Q L K H	192
Domain II	D S T T C F N A K G P L L K K	378

Fig. 3. Comparison of the 3 domains of Gc. Identical amino acid residues are boxed.

TABLE 1. HOMOMLOGY SUMMARY

Homology Block	Location (Nucleotide Residue)	DNA Homology (%)	Pa(0.25)	Amino Acid Homology (%)
I	286-441 vs 859-1011	65/156 (42)	8.75×10^{-5}	14/52 (27)
II	514-630 vs 1081-1197	53/120 (44)	9.30×10^{-5}	9/40 (23)
III	868-1008 vs 1429-1575	59/147 (40)	4.41×10^{-4}	11/49 (22)

its third domain was missing a carboxy-terminal sequence contained by albumin and α -fetoprotein (9). Intragenic homology of Gc was also analyzed using the ALIGN program. Domain I (residues 1-192) aligned with Domain II (residues 193-378) produced a score of 11.61 SD units. Portions of domains I (residues 1-79) and II (residues 193-269) were compared to Gc's partial domain III (residues 379-458) to give scores of 5.25 SD and 6.18 SD units, respectively. Alignment scores for the domains indicated a high probability of a common ancestor.

The relative age of the Gc gene is difficult to estimate. Homology data indicated that albumin and α -fetoprotein evolved from an ancestral gene that was produced by an intragenic triplication 300-500 million years ago (19). Two lines of evidence suggest that the Gc gene is older than the other two members of the albumin family. Albumin and α -fetoprotein are more similar to each other than either is to Gc, indicating that the gene duplications producing them may have occurred more recently. Furthermore, Gc, alone, contains four half-cystine residues at positions 13, 58, 59 and 67 that are capable of forming the double disulfide bond predicted by Brown (11) to have occurred in the first domain of the triplicated ancestral gene (9). Whether or not the partial loss of Gc's third domain is a recent evolutionary event must await further studies of the Gc gene in other species.

When the cDNA sequences (9) of the three Gc domains were compared by statistical analysis to identify regions of extensive internal nucleotide sequence identity (12), 3 paired sequences, designated as homology blocks I, II and III in Table 1, demonstrated homologous regions of at least 100 base pairs long and with a Pa of 7×10^{-4} or less. The 3 pairs of homologous sequences define the internal alignment, shown in Figure 4. Therefore, the Gc sequence can be divided into three segments that correspond to the three domains described in human albumin by Brown (11). This alignment places all of the Cys residues in phase. When the amino acid sequences within each

Fig. 4. Domain homology blocks. Horizontal lines represent the cDNA sequence of Gc divided into Domains I, II and III (partial domain) with vertical lines identifying statistically significant homology blocks. Dots symbolize cysteine residues. Numerals indicate nucleotide boundaries of each domain (amino acid residues in parentheses).

homology block are compared, the results are in accord with the DNA homology. As in the case of albumin, the first two domains of Gc are more similar than either is to the third domain.

The basis for evolutionary constraints on the codons contained in the three homology blocks of Gc is unknown. Diversity in the nucleotide sequence of domains I and II of the Gc gene (64%) can be compared to the lower rate of divergence seen in nucleotides encoding internal, homologous domains of two other human plasma protein genes: the human transferrin (Tf) gene, 50%, (10) and the human haptoglobin (Hp2) gene, 1% (20).

The secondary structure of human Gc, as predicted from the amino acid sequence according to the procedures of Chou and Fasman (13,14) is shown in Figure 5. A summary of the predictions of secondary structures of human albumin, α -fetoprotein and Gc2 is given in Table 2. The percentages of β -turns, β -sheets, α -helices and random coils were calculated. When a sequence could participate in more than one conformation it was included in more than one calculation. Our derivation of the secondary structure of albumin differs somewhat from that of Morinaga et al. (21) by including Chou and Fasman's (14) method of finding beta turns, by making greater use of boundary parameters through computer assistance, and by placing albumin's N-terminal dicysteine at positions 90-91 as was indicated by the amino acid sequence published by Dugaiczky et al. (22), instead of positions 89-90 as was indicated by the secondary structure of Morinaga et al. (21).

The predicted secondary structure of Gc2 differed mainly in the percentage of β -turns when compared to human albumin and α -fetoprotein, with a predicted 42% of its sequence participating in β -turns, compared to 26% in albumin and 34% in α -fetoprotein.

Fig. 5. Predicted protein secondary structures of human Gc2. The amino acid residues forming α -helices, β -sheets, β -turns and random coils are indicated by $\odot$, $\oplus$, $\ominus$ and $\circ$, respectively. Some residues participate in both α -helices and β -turns, or in both β -sheets and β -turns, and these are indicated by $\oplus$ and $\ominus$, respectively. The arrow indicates one sequence at which N-linked glycosylation may occur. The loops formed by disulfide bonding are filled in black.

TABLE 2. PREDICTED^a SECONDARY STRUCTURE OF ALBUMIN, Gc AND α -FETOPROTEIN

Protein	α -helix (%)	β -sheet (%)	β -turn (%)	Random Coil (%)
Human Albumin (22)	47.4	13.7	26.0	15.4
Human Gc2 (9)	29.9	14.0	42.0	17.5
Human α -Fetoprotein (21)	36.4	23.0	34.0	9.5

^a(13,14)

Gc2 and α -fetoprotein each have one potential N-glycosylation site in their second domains while albumin has none (22). The potential glycosylation site in Gc2 occurs at residues 272-274 (Asn-Leu-Ser) (9). Previous studies (3-5) demonstrated that Gc1, but not Gc2, has an O-glycosylation site within the sequence of residues 412-424, probably at the Thr residue at position 420 (4,9). Residue 420 is occupied by Lys in Gc2 (9).

Utilization of Gc cDNA has also led to the identification of its genomic

Fig. 6. Filter hybridization of human Gc2 cDNA with human lymphocyte DNA digested with *Pst*I (A), *Eco*RI (B) and *Pvu*II (C). A DNA polymorphism, 12 kbp, can be seen in the last lane of C.

DNA, and discovery of a DNA polymorphism detected with *PvuII*. Figure 6 is filter hybridization of radiolabeled *Gc* cDNA and endonuclease digests of human lymphocyte DNA. Six fragments of *PstI* digest (14.0, 9.9, 6.5, 5.4, 4.5 and 2.2 kbp) are seen in Figure 6A. In Figure 6B, 10 fragments appear in the *EcoRI* digest (10.0, 6.7, 4.5, 3.0, 2.9, 2.6, 2.5, 2.0, 1.9 and 1.7 kbp), and 5 fragments of *PvuII* digest (15.0, 8.4, 8.2, 5.3 and 5.0 kbp) were found to hybridize with *Gc* cDNA. A 12 kbp *PvuII* polymorphic fragment was detected in one out of 14 individuals (Figure 6C). These results together with the hybridization data obtained from *BglII*, *KpnI*, *TaqI*, and *BamHI* digests (data not shown) suggest that human *Gc* is encoded by approximately 18 to 24 kbp DNA. Further studies will be required to determine evolutionary relationships by comparison of the number and arrangement of introns among the *Gc*, albumin and α -fetoprotein genes. Intron comparison and arrangement in the genes encoding *Gc*, albumin and α -fetoprotein are presently being studied to trace the evolutionary basis for the partial deletion of the C-terminal domain of the *Gc* gene. Examination of *Gc* in other species may also be helpful in determining whether or not the partial deletion of the third domain is a recent or ancient event in the evolution of this gene. The detection of DNA polymorphisms with *Gc* cDNA may prove useful in prenatal diagnosis of dentinogenesis imperfecta, an autosomal dominant disease mapped on chromosome 4 approximately 7 centimorgans from the *Gc* gene (23).

ACKNOWLEDGEMENTS

We thank Judith Gardella for technical assistance, Betty Russell and Judith Pride for preparation of the manuscript. This work was supported in part by NIH grants HD16584, GM33298, and AM34992, a grant from the American Heart Association and American Cancer Society grant NP-470. R.D.M. was aided by Summer Science Research Program for Medical Students Grant No. 8-85-51 from the March of Dimes Birth Defects Foundation.

REFERENCES

1. Daiger, S.P., Schanfield, M.S. and Cavalli-Sforza, L.L. (1975) Proc. Natl. Acad. Sci. USA 72, 2076-2080.
2. Cleve, H. and Bearn, A.G. (1962) Prog. Med. Genet. 2, 64-82.
3. Svasti, J., Kurosky, A., Bennett, A. and Bowman, B.H. (1979) Biochem. 18, 1611-1617.
4. Svasti, J. and Bowman, B.H. (1978) Proc. Natl. Acad. Sci. USA 75, 4188-4194.
5. Coppenhaver, D.H., Sollenne, N.P. and Bowman, B.H. (1983) Arch. Biochem. Biophys. 226, 218-223.
6. Weitkamp, L.R., Rucknagel, D.L. and Gershowitz, H. (1966) Amer. J. Hum. Genet. 18, 559-571.

7. Harper, M.E. and Dugaiczky, A. (1983) *Am. J. Hum. Genet.* 35, 565-572.
8. Mikkelsen, M., Jacobsen, P. and Henningsen, K. (1977) *Hum. Hered.* 27, 105-107.
9. Yang, F., Brune, J.L., Naylor, S.L., Cupples, R., Naberhaus, K.H. and Bowman, B.H. (1985) *Proc. Natl. Acad. Sci. USA* (in print).
10. Yang, F., Lum, J.B., McGill, J.R., Moore, C.M., Naylor, S.L., van Bragt, P.H., Baldwin, W.D. and Bowman, B.H. (1984) *Proc. Natl. Acad. Sci. USA* 81, 2752-2756.
11. Brown, J.R. (1976) *Fed. Proc. Am. Soc. Exp. Biol.* 35, 2141-2144.
12. Sargent, T.D., Yang, M. and Bonner, J. (1981) *Proc. Natl. Acad. Sci. USA* 78, 243-246.
13. Chou, P.Y. and Fasman, G.D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.* 47, 243-246.
14. Chou, P.Y. and Fasman, G.D. (1979) *Biophys. J.* 26, 367-384.
15. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6-13.
16. Bell, G.I., Karem, J.H. and Rutter, W.J. (1981) *Proc. Natl. Acad. Sci. USA* 78, 5759-5763.
17. Barker, W.C. and Dayhoff, M.O. (1982) *Proc. Natl. Acad. Sci. USA* 79, 2836-2839.
18. Doolittle, R.F. (1981) *Science* 214, 149-159.
19. Eiferman, F.A., Young, P.R., Scott, R.W. and Tilghman, S.M. (1981) *Nature* 294, 713-718.
20. Yang, F., Brune, J.L., Baldwin, D., Barnett, D.R. and Bowman, B.H. (1983) *Proc. Natl. Acad. Sci. USA* 80, 5875-5879.
21. Morinaga, T., Sakai, M., Wegmann, T.G. and Tamaoki, T. (1983) *Proc. Natl. Acad. Sci. USA* 80, 4604-4608.
22. Dugaiczky, A., Law, S.W. and Dennison, O.E. (1982) *Proc. Natl. Acad. Sci. USA* 79, 71-75.
23. Ball, S.P., Cook, P.J.L., Mars, M. and Buckton, K.E. (1982) *Ann. Hum. Genet.* 46, 35-40.