Dissertations   **Theses and Dissertations**

1995

Cloning of Multiple Novel Human Trinucleotide Repeat Containing CDNA's: A Novel Application of Rapid Amplification of CDNA Ends (RACE)

James P. Carney
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Biochemistry Commons

LOYOLA UNIVERSITY CHICAGO

CLONING OF MULTIPLE NOVEL HUMAN TRINUCLEOTIDE REPEAT CONTAINING

cDNA'S: A NOVEL APPLICATION OF RAPID AMPLIFICATION OF cDNA ENDS

(RACE)

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL IN

CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MOLECULAR AND CELLULAR BIOCHEMISTRY

BY

JAMES P. CARNEY

CHICAGO, ILLINOIS

JANUARY, 1995

ACKNOWLEDGMENTS

DEDICATION

This work is dedicated to the memory of my parents George J. and Norene G. Carney and my sister Joyce N. Carney. They have provided me with a constant source of inspiration throughout my graduate career.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

LIST OF ABBREVIATIONS

μCi	microcurie
μg	microgram
μl	microliter
μM	micromolar
bp	base pair
Ci	curie
CIAP	calf intestinal alkaline phosphate
Da	daltons
DEPC	diethyl pyrocarbonate
DM	myotonic dystrophy
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DRPLA	dentatorubral palladoluysian atrophy
DTT	dithiolthreitol
EDTA	ethylenediamine tetraacetic acid
FRAXA	fragile X syndrome A
FRAXE	fragile X syndrome E
g	gram
GIT	guanidinium isothiocyanate

HD	Huntington's disease
kb	kilobases
kDa	kilodaltons
LB	Luria broth
M	molar
Mb	megabases
mg	milligram
ml	milliliter
mm	millimeter
mM	millimolar
mRNA	messenger RNA
O.D.	opitical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RNA	ribonucleic acid
SBMA	spino-bulbar muscular atrophy
SCA1	spinocerebellar ataxia type 1
SDS	sodium dodecyl sulfate
TBE	tris-boric acid-EDTA-electrophoresis buffer
TE	tris-EDTA buffer

TEMED	tetramethylethylenediamine
tRNA	transfer RNA
UDG	uracil-DNA glycosylase
UV	ultraviolet
X g	times gravity

ABSTRACT

CLONING OF MULTIPLE NOVEL HUMAN TRINUCLEOTIDE REPEAT CONTAINING

cDNA'S:

A NOVEL APPLICATION OF RAPID AMPLIFICATION OF cDNA ENDS (RACE)

The expansion of trinucleotide repeat sequences is a process by which the number of GC rich triplet repeats within a specific locus in the genome is amplified leading to a disease state. Presently, seven disorders have been shown to be the result of this type of mutation. These disorders are dentatorubral-palladoluysian atrophy (DRPLA), Fragile X syndrome(A) (FRAXA), Fragile X syndrome(E) (FRAXE), Huntington's disease (HD), myotonic dystrophy (DM), spino-bulbar muscular atrophy (SBMA), and spinocerebellar ataxia type 1 (SCA1). A subset of these disorders, DRPLA, HD, SBMA, and SCA1, are caused specifically by the expansion of unstable $(CAG)_N$ repeats located within translated regions of the respective transcripts and appear to define a subclass of trinucleotide repeat expansion disorders. I report here an initial step towards characterizing other disorders of this subclass. Utilizing rapid amplification of cDNA ends, I have isolated multiple novel human cDNA's that contain $(CAG)_N$ repeats.

These cDNA's should provide useful reagents for further investigation into trinucleotide repeat expansion disorders.

Initially using a 3' Rapid Amplification of cDNA Ends (RACE) approach I isolated six trinucleotide repeat containing cDNA's. From this group, two were the focus of further analysis. The cDNA sequence for clone CAG8-6 was determined and has no significant similarity with any sequences in GenBank. In collaboration the gene for CAG8-6 was mapped to chromosome 1q41-42. The cDNA from clone CAG8-16 was completely sequenced and by GenBank search was found to encode the human homologue to a previously characterized mouse single-stranded DNA binding protein (ssbp). The proteins from mouse and human showed a striking degree of conservation being over 90% identical. Somatic cell hybrid panel analysis indicates that the human ssbp maps to chromosome 5. Additionally, two known cDNA fragments were isolated which indicated the utility of the technique. The calmodulin 1 (*CALM1*) cDNA was one of these two known cDNA's. Previous to this work it was unknown that the *CALM1* gene contained a CAG repeat. The novel clones isolated should provide molecular probes for further investigation in to their possible involvement in disorders caused by trinucleotide repeat expansion.

The latter portion of the project focused on the development and utilization of a novel technique, which I call Random RACE. The 3' RACE technique has the inherent limitation that one can

only isolate trinucleotide repeat containing cDNA's which have their repeat located near the poly A⁺ tail. Random RACE allowed for the elimination of this limitation and the isolation of cDNA fragments from trinucleotide repeat containing transcripts regardless of the location of the repeat. Utilizing this technique, greater than 30 novel human cDNA fragments have been isolated. Genbank searches have indicated some regions of DNA sequence similarity in a number of the clones which may provide a basis for characterizing the function of these gene products. These clones constitute a molecular library that can be utilized for screening other genetic disorders that are caused by the expansion of CAG repeats.

CHAPTER I

INTRODUCTION

The onset of molecular biology has brought about a revolution in the life sciences. The techniques available have led to rapid advances resulting in greater understanding of cellular processes. Human molecular genetics has greatly benefited from these advances through better diagnosis and the possibility of improved treatment. The ability to analyze and manipulate the DNA molecule has led to better diagnosis of human diseases and with the onset of gene therapy, a new age of improved treatment is upon us. The application of molecular biology techniques to diagnosis and prognosis of human disease will bring a plethora of discoveries providing a greater understanding of numerous human genetic diseases.

One of the most exciting recent findings of human molecular genetics is the occurrence of trinucleotide repeat expansion in human disease states. The expansion of GC rich trinucleotide repeat sequences in DNA is now known to be a major type of mutagenesis leading to human disease states (Richards and Sutherland, 1992). In the last three years seven diseases have been described that are caused by trinucleotide repeat expansion.

Presently, these disorders appear to segregate into two groups. One group, resulting in four different dominant-late-onset neurological disorders, is caused by expansion of an unstable $(CAG)_N$ repeat that is located in a translated region of the respective genes (LaSpada et al., 1991; HD Collaborative Research Group, 1993; Orr et al., 1993; Koide et al., 1994; Nagafuchi et al., 1994). The four disorders of this group are spinobulbar muscular atrophy (SBMA), Huntington's disease (HD), spinocerebellar ataxia type 1 (SCA1), and dentatorubralpallidoluysian atrophy (DRPLA). In all cases the $(CAG)_N$ repeat is translated as polyglutamine. With the exception of SBMA, the cellular function of all of the respective gene products is unknown. Given the dominant nature of the disorders one possible molecular mechanism is that the protein products are involved in some novel interaction as a result of the expanded polyglutamine region (HD Collaborative Research Group, 1993; Orr et al., 1993). Such regions of polyglutamine are known to be important in a number of transcription factors (Gerber et al., 1994) and may have a role in the evolution of protein sequences (Green and Wang, 1994).

The second group of disorders is caused by the expansion of GC rich trinucleotide repeat located in untranslated region of the three respective genes (Fu et al., 1991, Brook et al., 1991, Knight et al., 1993). The disorders of this group are Fragile X syndrome A (FRAXA), Fragile X syndrome E (FRAXE), and myotonic

dystrophy (DM). The expansion in these disorders apparently affects the transcript levels of the respective genes, however, the mechanism of the resulting pathology is largely unknown.

Considering that these seven disorders have been described in such a short amount of time there is a widely held tenet that there exist a number of other disorders that are caused by the expansion of trinucleotide repeats (Richards and Sutherland, 1992; Caskey and Kuhl, 1993). It is known that there are a number of diseases whose molecular cause is presently undefined that have characteristics similar to the seven disorders now known. In particular, several neurodegenerative ataxias have been described that show genetic anticipation and clinical variability (H. Zoghbi, personal communication). These characteristics suggest that these disorders are caused by trinucleotide repeat expansion. It is the aim of my dissertation to utilize Rapid Amplification of cDNA Ends (RACE) to isolate $(CAG)_N$ containing cDNA's. These cDNA clones will comprise a useful molecular database for screening genetic disorders suspected to be caused by expansion of trinucleotide repeats.

I have utilized two separate RACE applications to accomplish the isolation of $(CAG)_N$ containing cDNA's.

1. 3' RACE

Using reverse transcribed RNA as a template this technology allows for amplification between the poly A^+ tail of an mRNA and a unique internal sequence. In this work this internal sequence was a $(CAG)_N$ containing primer. This adaptation allows for the amplification and cloning of any cDNA that contains a CAG repeat located within approximately one kilobase of the poly A+ tail. $(CAG)_N$ was chosen as the primer sequence due to the involvement of CAG repeats in the translated regions of the genes that are defective in the group 1 disorders. The subsequent fragments isolated are sequenced, utilized as probes for expression analysis, cDNA cloning, and chromosomal localization.

2. Random RACE

This methodology represents a novel application of RACE developed for this work. The technique allows for the amplification between a unique known sequence and a random sequence present in a cDNA. As in 3' RACE, the unique sequence primer contains a $(CAG)_N$ region. In this adaptation the method is utilized to clone $(CAG)_N$ containing cDNA fragments regardless of where the repeat is located within a mRNA. The method overcomes the limitation of 3' RACE, which requires the repeat to be located within a reasonable distance of the poly A^+ tail. The DNA

sequence of the isolated fragments will be determined and novel clones identified by searching GenBank.

The significance of carrying out the above studies is that the isolation of multiple novel CAG repeat containing cDNA fragments will generate a molecular library that can be utilized in future experiments aimed at the molecular dissection of human diseases caused by the expansion of $(CAG)_N$ sequences. Given the suspicion that a large number of human diseases are caused by the expansion of $(CAG)_N$ sequences it is not unreasonable to expect that the reagents generated by this work will prove to be useful in future studies.

CHAPTER II

REVIEW OF RELATED LITERATURE

A. Genome Dynamics

Many examples of genome alterations are known, including gene amplification and loss of heterozygosity that occur in cancer cells (for review, see Cheng and Loeb, 1993). For example, instability at microsatellite repeats in hereditary non-polypsis colon cancer has been shown to be due to defects in DNA mismatch repair (Aaltonen et al., 1993; Fishel et al., 1993; Leach et al., 1993; Thibodeau et al., 1993; Bronner et al., 1994; Papadopoulos et al., 1994). One important type of dynamic genomic element is the variable nucleotide tandem repeat (VNTR) or microsatellite repeat. This element consists of a sequence of one to six bases that can be repeated multiple times (Tautz, 1989). The most prevalent type of VNTR is the dinucleotide repeat. This repetitive element has become very important as a tool for genetic mapping. Dinucleotide repeats offer the advantage of being highly polymorphic (variable repeat numbers at the same locus within the population) and are within a size range to allow for PCR amplification and accurate size determination (Tautz, 1989; Weber

and May, 1989). Using this methodology it is now possible to perform linkage analysis with high resolution in a relatively short period of time. Additionally, this method has been utilized as one of a battery of techniques to create a first generation physical map of the human genome (Cohen *et al.*, 1994). Along with dinucleotide repeats, tri- and tetra-nucleotide repeats are other highly polymorphic VNTR's. These repetitive elements also offer the advantages of small size to allow size determination by PCR (Edwards *et al.*, 1991). The only disadvantage of these repetitive elements is they are in lower abundance on a genome-wide basis. These repetitive elements are useful for both genetic mapping purposes and forensic identification. Through the use of five different loci Edwards *et al.* (1991) have shown that one can match an individual with only a 1 in 90,000 chance of having a random match. If the number of loci is increased to twelve the odds of a random match increases to 1 in 1 X 10^8. This method has the advantage of being PCR based and therefore it does not require a large sample. This makes the technique well suited to forensic applications.

Trinucleotide repeats are useful in both genetic mapping and forensic analysis, yet they have become an area of intense research due to their involvement in a number inherited diseases. Trinucleotide repeats have been observed in a large number of genes from a variety of species (Grabowski *et al.*, 1991; Gerber *et al.*, 1994). One type of repeat, $(CAX)_N$ where X= A, C, or G, was

originally described in the *Notch* gene of *Drosophila* and termed *opa* or *strep* (Wharton et al, 1985). A similar repeat, $(CAG)_N$, is found in a large number of human genes (Table I) and is expanded in four late onset neurological diseases (see below). A definitive function for this repeat is not known, however, it is often present in translated regions of genes and has a high propensity to code for glutamine (Han *et al.*, 1994; Stallings, 1994). Table I shows that a large number of the genes that contain a $(CAG)_N$ trinucleotide repeat code for transcription factors or other cellular control proteins. Functionally, these polyglutamine stretches are important for the protein-protein interactions (Gerber *et al.*, 1994) necessary for transcriptional activation. However, analysis using transient transfections of constructs containing a polyglutamine region fused to the DNA binding domain of the yeast transcriptional activator *GAL4* has shown that transcriptional activation reaches maximal level with about 30 glutamines (Gerber *et al.*, 1994). This is similar to the upper limit of repeats observed in the normal population for the genes involved in the four late onset neurological disorders (Table II).

The $(CAG)_N$ repeat has been shown to be one of the most prevalent repetitive elements in human GenBank DNA (Green and Wang, 1994; Han et al., 1994; Stallings, 1994) and is the most abundant repeat present in human exonic sequences (Stallings, 1994). Additionally, Stallings (1994) has observed that frequently

TABLE 1:   HUMAN GENES CONTAINING $(CAG)_N$ REPEATS

Gene	[a]Repeats	Accession Number
TBP	23	X54993
*Androgen Receptor	13-30	J03180
*Huntingtin	11-34	L12392
*Ataxin-1	6-39	X79204
*CTG-B37	7-25	L10377
MEF-2	11	S43912
IL-9 Receptor	10	M84747
RSRFC4/9	9	X63381
Serum Response Factor	9	S70452
Pim-1 proto-oncogene	8	M27903

Table 1:   The table shows a partial listing of genes containing $(CAG)_N$ repeats.   The table was generated by searching GenBank with the sequence $(CAG)_{10}$.   All of the entries in the table have at least 8 identical repeats of CAG.   [a]Number of repeat units. *These entries are genes that have been implicated in diseases caused by expansion of trinucleotide repeats (see Table 2).

trinucleotide repeats located within translated regions are not conserved indicating that tracts of certain poly amino acids, in particular polyglutamine, are not critical for protein function. Consistent with this data, Green and Wang (1994) have proposed that insertion of polyglutamine tracts within protein sequences is an evolutionary mechanism that allows proteins to add amino acids. The next step in this process would be base substitution mutations which would alter the repeat and, under selective pressure, could create new protein domains (Green and Wang, 1994).

B. **Trinucleotide Repeat Expansion**

Trinucleotide repeat expansion is a type of mutagenesis where, in general, the mutation frequency of the repetitive element is based upon its size (Richards and Sutherland, 1992) leading to the term **dynamic mutation** to describe this process. The mutagenesis observed is an increase (or decrease) in the number of repeats present within a gene. Figure 1 illustrates this process in qualitative terms, showing the dependence of mutation frequency on the repeat size. The result of this phenomena is the general increase in repeat size in an affected family over generations. This increase in repeat size correlates with severity of the phenotype and is termed anticipation. Additionally, some of the disorders show a correlation between age of onset and repeat length (Brook *et al.*, 1992; Fu *et al.*, 1992; Tsilfidis *et al.*, 1992; HD Collaborative Research Group, 1993; Orr *et al.*, 1993; Koide *et al.*, 1994).

The mechanism of trinucleotide repeat expansion remains undetermined. The most often hypothesized mechanism is that of replicative slippage followed by ineffective mismatch repair. This model is supported by experiments in *Escherichia coli* which showed that defects in the mismatch repair genes *mutL* and *mutS* genes lead to a 13 fold elevated level of repeat tract instability (Levinson and Gutman, 1987). Also recent work in yeast has shown that defects in the mismatch repair genes *PMS1*, *MLH1*, and *MSH2* lead to a 100 to 700 fold increase in instability of microsatellite repeats (Strand et al., 1993). Strand *et al.* (1993) also showed that mutations in the proofreading activities of DNA polymerase δ and DNA polymerase ϵ had little effect on dinucleotide tract instability. These results would seem to indicate that the level of polymerization slippage is normally at a near maximal level but that these slippage errors are efficiently corrected by the mismatch repair system.

Finally, work on hereditary nonpolypsis colon cancer (HNPCC) has shown that defects in mismatch repair are responsible for tumorigenesis. HNPCC cells exhibit genome wide instability in microsatellite repeats (Altonen et al., 1993; Thibodeau et al., 1993) From linkage analysis of affected families, two different mismatch repair defects were localized to chromosome 2 and chromosome 3 (Peltomäki et al., 1993; Lindblom et al., 1993). The gene *hMSH2* was cloned and localized to chromosome 2 by two groups and shown to be mutated in HNPCC affected families(Fishel et al., 1993; Leach et al., 1993). The predicted amino acid sequence of the human *MSH2* shows 77% identity to the yeast *MSH2* (Fishel et
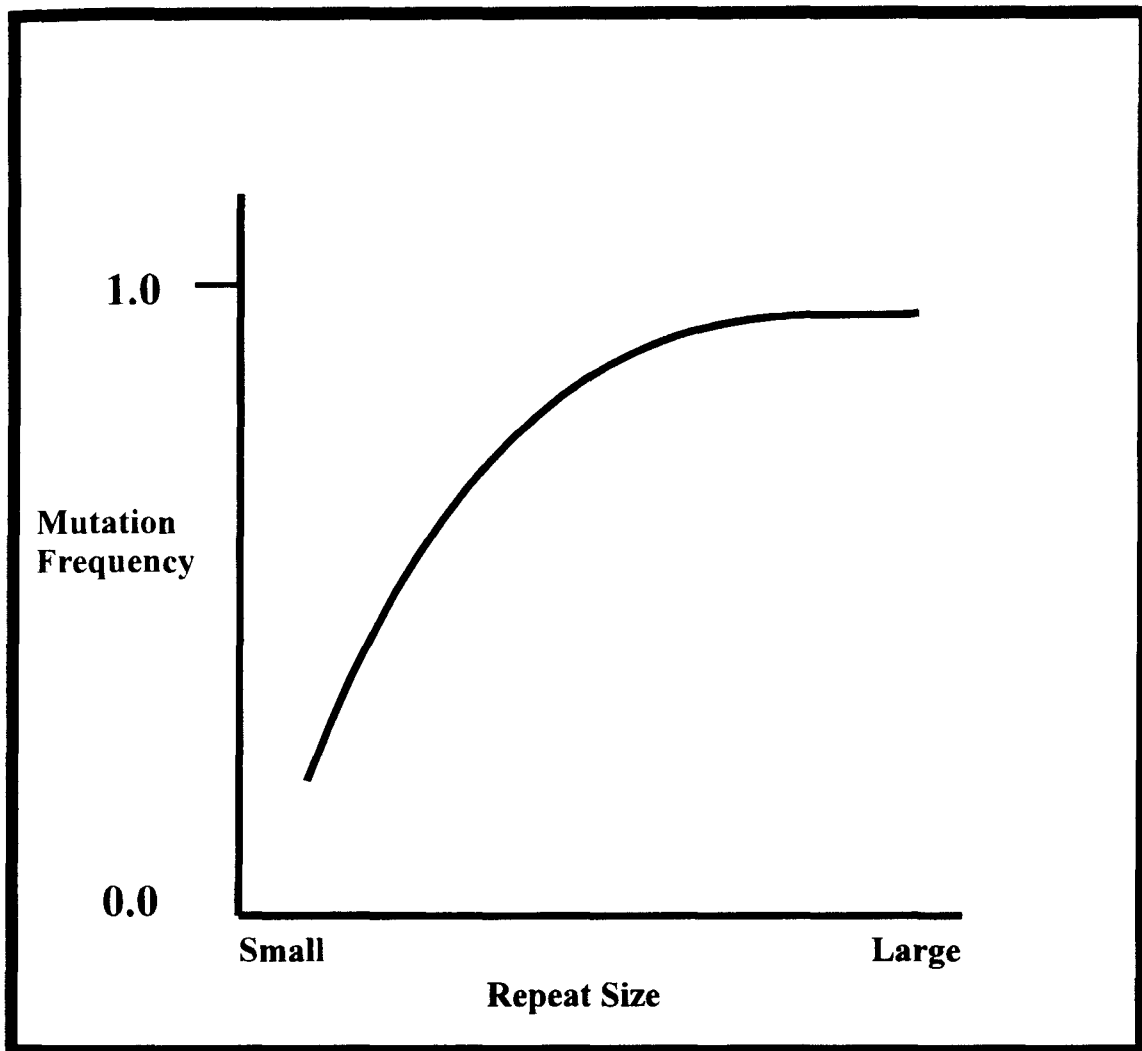
Figure 1: The graph gives an approximate representation of dynamic mutation with the mutation frequency being dependent on repeat size. The dynamic nature of repeat mutatgenesis appears to be dependent on many factors and is not thought to apply to normal alleles (Adapted from Richards and Sutherland, 1992 and Kuhl and Caskey, 1993).

al., 1993). The gene *hMLH1* was cloned and shown to reside on chromosome 3 and was mutated in HNPCC families (Bronner *et al.*, 1994; Papadopoulos *et al.*, 1994). The ORF for the human MLH1 shows 34% identity to the yeast MLH1 (Papadopoulos et al., 1994). This data indicates that in humans defects in the mismatch repair system lead to instability in VNTR's and predisposition to cancer.

C. Late Onset Neurological Disorders

1. Spinobulbar Muscular Atrophy (SBMA)

SBMA is a rare X-linked recessive disorder originally described by Kennedy and coworkers and sometimes referred to as Kennedy's disease (1968). The disorder is clinically characterized by onset in the third to fifth decade of life followed by progressive muscle weakening and atrophy (Harding *et al.*, 1982). Symptoms include muscle cramps that precede onset of the disease by several years, facial weakness, fasciculation, and gynaecomastia. The presence of gynaecomastia led to the hypothesis that the disease was caused by a mutation that created an endocrine defect.

Linkage analysis showed the disease to be linked to markers on the X chromosome in the same region of the androgen receptor (Fishbeck *et al.*, 1986). LaSpada *et al.* (1991) reported that the gene defect was an increase in the number of CAG repeats present

in the first exon of the androgen receptor gene (Table 2). An analysis of 75 controls showed the repeat to be polymorphic in the population with an average repeat length of 21±2, with a range of 13-30. The expanded disease allele showed an absolute association with the disease with a range of 40-62 repeats in affected patients(LaSpada *et al.*, 1991). The $(CAG)_N$ repeat begins at codon 58 of the androgen receptor protein and codes for a polyglutamine tract (Lubahn *et al.*, 1988). Mhatre *et al.*(1993) have shown in transient transfection transcription assays that an androgen receptor with an expanded polyglutamine tract suboptimally transactivates a reporter construct carrying four copies of the androgen response element. This is in agreement with the work of Gerber *et al.* (1994), discussed above, who used an artificial system to show that large (>30) polyglutamine stretches did not transactivate as effectively as tracts <30 units in length. Additionally, Warner *et al.* (1991) have shown that fibroblasts from SBMA patients have decreased androgen binding capability. This data would seem to indicate a loss of function mutation, however, Fishbeck argues an alternative possibility. It is hypothesized that the increase in the polyglutamine tract of the androgen receptor results in a toxic gain of function. It is proposed that the androgen receptor carrying a larger polyglutamine tract either interacts aberrantly with its protein targets or interacts with a new protein target. The absence of a

TABLE 2: DISEASES OF TRINUCLEOTIDE REPEAT EXPANSION

A. LATE ONSET NEUROLOGICAL DISORDERS

Disease	[a]Repeat	[b]Normal Range	[b]Disease Range	Location	Chromosome
SBMA	CAG	13-30	40-62	coding	Xq11-12
HD	CAG	11-34	38-100	coding	4p16.3
SCA1	CAG	6-39	43-81	coding	6p22-23
DRPLA	CAG	7-25	49-75	coding	12p12-ter

B. OTHER DISORDERS OF TRINUCLEOTIDE REPEAT EXPANSION

Disease	[a]Repeat	[b]Normal Range	[b]Disease Range	Location	Chromosome
DM	CTG	5-30	50->200	3'-UTR	19q13.3
FRAXA	CGG	6-50	>200	5'-UTR	Xq27.3
FRAXE	GCC	6-25	200->700	?	Xq27-28

Table 2: The table summarizes the characteristics of the genes implicated in diseases of trinucleotide repeat expansion. Section A includes the late onset neurological disorders where the $(CAG)_N$ repeat is located in the translated region of the four respective genes. Section B is made up of other diseases of trinucleotide repeat expansion where the GC rich repeats are present in untranslated regions of the respective genes. [a]The repeat is given as it reads on the coding strand. [b]The ranges for both normal and disease alleles are given in repeat units.

phenotype in females carrying an expanded allele is then explained on the basis of Lyonization and lower androgen levels in females (K. Fishbeck, personal communication). In support of this hypothesis Biancalana et al. (1992) have reported heterozygous carrier females that have complained of muscle cramps, indicating a possible mild expression of the disease phenotype. Biancalana et al. (1992) also reported that the disease allele in SBMA shows only moderate instability. The authors examined a four generation family affected by SBMA and found that the most the repeat expanded from one generation to the next was 5 units. Additionally, there has been no report of mitotic instability or mosaicism in SBMA affected patients. This is in contrast to a number of other trinucleotide repeat expansion disorders that often show large increases from one generation to the next. Overall, the influence of the expanded polyglutamine tract on the pathophysiology of the disease is presently unclear and will require further study.

2. **Huntington's Disease (HD)**

HD is an autosomal dominant disorder with an incidence of 1 in 10,000 with onset generally in the third to fifth decade of life. However, juvenile onset cases have been reported and these typically show more severe symptoms and a faster progression (Gusella et al., 1993). In addition, juvenile onset HD is generally associated with paternal transmission of the disease (Telenius et al., 1993). The span of the lethal disease from the

onset of symptoms is approximately 20 years. Clinically the disorder is characterized by motor disorders (chorea), cognitive loss, and personality disorders (Martin and Gusella, 1986). The neuropathology of HD displays selective loss of neurons mainly in the caudate nucleus and putamen (Gusella *et al.*, 1993).

The underlying genetic defect in HD was mapped to chromosome 4p in 1983 (Gusella *et al.*). The focus of the following ten years of research was to isolate the defective gene with this search ending in 1993. Through the use of exon trapping, exons were isolated from the HD candidate region at 4p16.3 and several were found to correspond to a transcript called IT15 (HD Collaborative Research Group, 1993). A $(CAG)_N$ repeat present in the 5' region of this transcript was found to be expanded and unstable in disease pedigrees. This repeat falls within the predicted reading frame and codes for a polyglutamine tract. The repeat is polymorphic in the normal population showing a range of 11 to 34 repeat units while disease alleles show a range of 38 to 100 repeat units (HD Collaborative Research Group, 1993).

Analysis of the HD gene has demonstrated that it is made up of 67 exons spread out over 185 kb with the repeat located in exon 1 (Ambrose *et al.*, 1994). The gene generates two transcripts of 13.5 and 10 kb which Ambrose *et al.*(1994) have reported differ by alternative polyadenylation. However, Lin *et al.* (1994) have shown by PCR the existence of two alternatively spliced transcripts which differ by 1.4 kb and would correspond to a 480 amino acid region that would be absent in an isoform of the protein. The larger protein product is a 3,130 amino acid

polypeptide which contains a leucine zipper motif but no other similarity to any known genes (Hoogeveen et al., 1993).

The HD gene is expressed in neuronal cells of the dentate gyrus, hippocampus, and cerebellum (Strong et al., 1993). Additionally, Strong et al. (1993) have demonstrated expression in a variety of non-neuronal tissues including colon, liver, pancreas, and testes. Hoogeveen et al. (1993) have utilized immunocytochemistry to demonstrate the presence of the huntingtin protein in the cytoplasm of many cell types but with additional protein present in the nucleus of neuronal cell types. Furthermore, an interesting caveat to the neuropathology of the disorder is the observation by Telenius et al. (1994) of somatic mosaicism in HD patients, with the highest degree of expansion being present in the tissues that are most severely affected. Aside from this observation neither the expression pattern nor the subcellular location of the huntingtin protein offer any clue as to the molecular pathology of the disease.

Interestingly, several cases of sporadically occurring HD have been reported and appear to arise from expansion of large normal alleles in the range of 30 to 38 repeats (Goldberg et al., 1993; Myers et al., 1993). It seems that, similar to Fragile X syndrome, normal alleles with 35 to 40 repeats may be predisposed to expansion and thus constitute a premutation range. Both Goldberg et al. (1993) and Myers et al. (1993) hypothesize that cis acting elements on the disease chromosome may contribute to instability and the progression to a full HD mutation although the disease alleles have been shown to be associated with a number of

different haplotypes (MacDonald *et al.*, 1992) . Goldberg *et al.*
(1993) also reported that all of the sporadic cases studied in
their pedigrees occurred by expansion of a paternal premutation
indicating sex influence on HD instability. In support of this
Telenius *et al.* (1994) have shown a high degree of mosaicism in
sperm from affected males, indicating that expansion occurs during
spermatogenesis.

HD being a dominant disorder is expected to result from a
gain of function mutation and this hypothesis has been supported
by the observation that a patient with a balanced translocation
within the HD gene does not result in an HD phenotype (Ambrose *et
al.*, 1994) . Ambrose *et al.* (1994) have also shown that the
disease allele is expressed and therefore conclude that the
mutation confers a new property on the HD transcript or more
likely the protein. The exact nature of this altered property
will require further study that should provide characterization of
proteins that interact with the huntingtin polypeptide.

3. <u>Spinocerebellar ataxia type 1 (SCA1)</u>

Spinocerebellar ataxia type 1 is an autosomal dominant
neurodegenerative disorder that maps to the short arm of
chromosome 6 (Zoghbi *et al.*, 1988; Bryer *et al.*, 1992).
Clinically the disorder is characterized by ataxia,
opthalmoparesis, and motor weakness (Currier *et al.*, 1972). The
onset of symptoms generally occurs in the third or fourth decade
of life with a 10 to 20 year progression to death (Zoghbi *et al.*,

1988). Juvenile onset cases have been observed and they generally are more severe and show a faster progression to death with the disease allele usually inherited from an affected father. Additionally, anticipation is observed in families with a gradual decrease in the age of onset and the severity of symptoms through successive generations(Zoghbi *et al.*, 1988). Neuropathological analysis indicates selective neuron loss in the cerebellum and brain stem with degeneration of the spinocerebellar tracts (Greenfield, 1954). There is no biochemical defect known to be responsible for the neuronal loss.

The gene for SCA1 was originally mapped to chromosome 6 by linkage to the HLA locus (Zoghbi *et al.*, 1988). Further work localized the gene to a 1.2 Mb region flanked by the markers D6S274 and D6S89 and this region was cloned into four overlapping YAC's (Banfi *et al.*, 1993). Knowing the involvement of trinucleotide repeats in other disorders with similar characteristics to SCA1, Orr et al., (1993) screened the four YAC clones covering the region with trinucleotide repeat containing oligos. This allowed for the cloning of a fragment of the SCA1 gene which contained a polymorphic $(CAG)_N$ repeat that was expanded in affected individuals (Orr *et al.*, 1993). Initial analysis indicated that this repeat region was transcribed and Northern blot analysis with the cloned fragment detected an ~10 kb transcript. In addition, Orr *et al.* (1993)showed that the number of repeats on normal chromosomes was in a range of 6 to 39 while disease chromosomes have a range of 43 to 81 repeats with a strong correlation ($r = -0.845$) between repeat size and age of onset.

The association of juvenile onset SCA1 with paternal inheritance has been investigated and it has been observed that nearly 70% of maternal transmissions of the disease allele show no change in repeat size while 63% of paternal transmissions show an increase in the number of CAG repeats (Chung et al., 1993). Furthermore, Chung et al. (1993) showed by sequence analysis that 98% of normal SCA1 $(CAG)_N$ repeats are interrupted with at least one CAT while all expanded alleles are made up of pure $(CAG)_N$ repeats. This has led to the suggestion that loss of CAT interruptions in normal alleles may be a predisposing event to trinucleotide repeat expansion in SCA1 (Chung et al., 1993).

The SCA1 gene has been isolated and it has been shown to be made up of nine exons spanning 450 kb that generates a 10,660 bp transcript (Banfi et al., 1994). The first seven exons make up the 5' untranslated region while the last two contain the coding region and a 7,277 bp 3' untranslated region. The predicted reading frame generates a 816 amino acid, 87 kDa protein designated ataxin-1 (Banfi et al., 1994). DNA and amino acid sequence searches have revealed no significant similarity between ataxin-1 and any entries in a number of databases. Presently, the cloning of the gene that is defective in SCA1 offers little hint as to the molecular pathology of the disorder but does offer a rapid accurate method for diagnosis.

4. Dentatorubral-Pallidoluysian Atrophy (DRPLA)

DRPLA is an autosomal dominant neurodegenerative disorder that maps to the short arm of chromosome 12 (Koide *et al.*, 1994; Nagafuchi *et al.*, 1994). The clinical symptoms of DRPLA show a high degree of variability including cerebellar ataxia, movement disorders, and dementia (Naito *et al.*, 1982). Additionally, myoclonus epilepsy is also observed in some cases, usually those of juvenile onset (Takahashi *et al.*, 1988). Neuropathologic analysis revealed degeneration of the dentatorubral and pallidoluysian systems in all cases with heterogeneously occurring degeneration of the striatum and cerebellar cortex observed (Takahashi *et al.*, 1988). DRPLA is rare in populations of European descent yet shows increased incidence in the Japanese population. Additionally, DRPLA has recently been reported in an African-American family where it was originally named Haw River Syndrome, after the region of North Carolina where the affected family lives (Burke *et al.*, 1994). Similar to HD and SCA1 there is no known biochemical defect associated with DRPLA.

The gene that contains an expanded $(CAG)_N$ repeat that causes DRPLA was originally cloned by Li et al. (1993) by screening a cDNA library with poly CAG containing oligonucleotides. Li and coworkers (1993) isolated a number of trinucleotide repeat containing cDNA's, mapped them to human chromosomes, and investigated the polymorphic nature of several of the clones. Both Koide *et al.* (1994) and Nagafuchi *et al.* (1994) investigated the possibility of trinucleotide repeat expansion being the cause

of DRPLA by examining a number of the cDNA's isolated by Li et al. (1993). Both groups found that the $(CAG)_N$ repeat located within clone CTG-B37 is polymorphic in the population with a range of 7 to 25 repeats (Koide *et al.*, 1994; Nagafuchi *et al.*, 1994). Affected patients have one allele in the normal range and a single expanded allele which is in the range of 49 to 75 repeats (Koide *et al.*, 1994; Nagafuchi *et al.*, 1994). Furthermore, Koide *et al.* (1994) have shown a correlation between age of onset and number of repeats ($r = -0.7$). Similar to HD and SCA1 preliminary analysis has shown that paternal inheritance of an expanded allele results in an increased expansion while maternal inheritance shows a decrease in the number of repeats (Koide *et al.*, 1994).

Presently, analysis of the DRPLA gene is incomplete. Nagafuchi *et al.* (1994) have stated that the DRPLA gene produces a 4.5 kb transcript although the tissue distribution of the gene expression is unpublished. Furthermore, there is no information on the gene structure or the predicted protein product. Perhaps this information will assist in the elucidation of the molecular mechanism of neuronal degeneration in DRPLA.

Overall, it is the involvement of $(CAG)_N$ sequences in these late onset neurological diseases that has led to the focus of this work being the cloning of cDNA fragments that contain this repeat. It is hypothesized that the novel clones described here will provide useful reagents for the examination of the molecular defect in other late onset neurological disorders.

D. **Other Disorders of Trinucleotide Repeat Expansion**

1. **Myotonic Dystrophy**

Myotonic dystrophy is an autosomal dominant disease that maps to the long arm of chromosome 19 (Whitehead *et al.*, 1982; Brook *et al.*, 1992). The disorder is the most common form of adult muscular dystrophy and is clinically characterized by myotonia and muscle weakness and wasting. In addition patients often exhibit a variety of other symptoms including cardiac conduction effects, smooth muscle defects, hypersomnia, cataracts, abnormal glucose response, and, in males, premature balding and testicular atrophy (Harper, 1989). The disorder shows a high degree of clinical variability both within and between families which has led to the classification of three different subgroups of affected patients. The first group is the mildest form that is observed in middle or old age and is characterized by cataract with little muscle defect. The classic form of the disease is characterized by myotonia and muscle weakness and generally has an age of onset in adolescence or early adulthood. The most severe form of the disease occurs congenitally and is associated with mental retardation (Harper and Dyken, 1972). Additionally, early pedigree analysis by Fleischer (1918) led to the hypothesis of anticipation in DM. This is a progressive worsening of the disease phenotype through successive generations. Although this hypothesis was usually rebutted by ascertainment bias, it was eventually shown to be true (Höweler et al., 1989).

The gene defective in DM was mapped to chromosome 19 in 1982 (Whitehead et al.) and further genetic and physical refinements (Brook et al., 1992 and references therein) led to the cloning of a region of DNA that contained an unstable $(CTG)_N$ trinucleotide repeat that was expanded in affected patients (Brook et al., 1992; Fu et al., 1992; Harley et al., 1992; Mahadevan et al., 1992). Analysis of the trinucleotide repeat in the normal population by PCR showed it to be polymorphic with a range of 5 to 30 repeats with over 50% of the alleles being 5 or 13 repeats (Brook et al., 1992; Fu et al., 1992; Mahadevan et al., 1992). Analysis of affected patients invariably showed one allele within the normal range and a second allele either missing, due to the inability of the PCR to amplify across the expanded repeat, or alleles greater than 50 repeats (Brook et al., 1992; Fu et al., 1992). Furthermore, analysis of affected families showed that the size of the expansion increased through successive generations and seemed to correlate with the age of onset and clinical severity (Brook et al., 1992; Fu et al., 1992; Tsilfidis et al., 1992). This observation supplies a molecular explanation to the anticipation previously observed in DM. Furthermore, it has been shown that a disease allele can expand either during paternal or maternal transmission yet the large expansions that lead to congenital DM appear to come exclusively from maternal transmission (Tsilfidis et al., 1992; Lavedan et al., 1993)

The cloning of the DNA region containing the unstable repeat in DM led to the observation that this repeat was present within a transcriptional unit and detects a 3.3 kb transcript on Northern

blots (Brook et al., 1992; Fu et al., 1992). The cDNA for the DM gene was subsequently cloned and sequenced and found by sequence comparison to encode a putative protein kinase named myotonin protein kinase (M-PK) (Brook *et al.*, 1992; Fu *et al.*, 1992). Also, sequence analysis revealed that the $(CTG)_N$ repeat was located within the 3' UTR of the M-PK cDNA (Brook *et al.*, 1992; Fu *et al.*, 1992). Preliminary analysis of the M-PK protein indicates that it phosphorylates tyrosine residues but lack of critical experimental controls leave the exact function of M-PK an unresolved issue at this point (Etongué-Mayer *et al.*, 1994). Furthermore, this is a surprising result considering sequence comparison indicated that M-PK is a member of the serine/threonine family of kinases (Brook *et al.*, 1992).

The molecular pathology of DM is difficult to understand as it is hypothesized a dosage effect is responsible for the disease phenotype and that alterations in expression levels of the expanded M-PK allele are responsible for the disease (Fu *et al.*, 1993; Novelli *et al.*, 1993; Sabouri *et al.*, 1993). However, two groups report that the expansion of the $(CTG)_N$ repeat in the 3' UTR of the M-PK gene results in a specific decrease in the steady state level of M-PK mRNA transcribed from the disease allele in adult tissues(Fu *et al.*, 1993; Novelli *et al.*, 1993) while a third group has shown that the expansion leads to an increase in the steady state M-PK mRNA levels in congenital patients (Sabouri *et al.*, 1993). It is hypothesized that differing mechanisms are in operation in congenital and classic adult onset DM (Sabouri *et al.*, 1993). However, in further support of the loss of expression

model, Carango *et al.* (1994) have shown that a DM cell line that has had the normal M-PK allele deleted has no detectable expression from the disease allele. Additionally, Carango *et al.* (1994) have shown that the M-PK transcript appears to accumulate in an unprocessed form indicating that the defect may lead to a reduction in processed transcript levels. In regards to a possible mechanism for the loss of expression, Shaw *et al.* (1993) have shown that there is no detectable alteration in methylation status at the DM locus while Wang *et al.* (1994) have demonstrated that oligos of $(CTG)_N$ show an increased efficiency in nucleosome assembly as the repeat size is increased. Wang *et al.* (1994) hypothesize that increased nucleosome assembly at the 3' region of the DM locus leads to transcriptional repression. The validity of this hypothesis will require further investigation.

Interestingly, DM is the only one of the trinucleotide repeat expansion disorders where contraction of expanded alleles has been documented (Ashizawa *et al.*, 1994). Brunner *et al.* (1993) reported two cases where by haplotype analysis offspring had inherited an abnormal chromosome from an affected parent but the $(CTG)_N$ repeat had contracted into the normal range. The authors investigated possible germ line mosaicism of the affected parents to explain the contraction however they found no repeats in the size range that were observed in the offspring. The authors discussed a possible gene conversion mechanism although one of the cases did not show the same number of repeats on the abnormal chromosome as on her father's normal chromosome. Indicating a possible direct reversal of the expansion mutation. Additionally,

O'Hoy *et al.* (1993) reported a case where haplotype analysis indicated an affected father passed an abnormal chromosome 19 on to his daughter yet analysis of the $(CTG)_N$ repeat size in the daughter showed only 13 repeats on the paternally derived chromosome. More detailed haplotype analysis revealed two tracts of DNA over a 7.2 kb region which were derived from normal paternal chromosome yet interrupted with two markers which are on the paternal disease chromosome. The authors proposed a discontinuous gene conversion event although they did not rule out reciprocal crossover. These contraction events are an interesting phenomena that to date have only been observed and DM and warrant further investigation.

2. **Fragile X syndrome A (FRAXA)**

Fragile X syndrome A (FRAXA) is an X-linked dominant disorder with incomplete penetrance that is the most common form of familial mental retardation (Gustavson *et al.*, 1986). Additionally, macroorchidism and a distinctive facies are often observed in affected males (Nussbaum and Ledbetter, 1990). It has been observed that 30% of carrier females show symptoms of mental retardation while 20% of males who carry a Fragile X chromosome are phenotypically normal (Sherman *et al.*, 1984). Members of this group are referred to as non-transmitting males and their daughters who receive the disease allele are unaffected but grandsons who subsequently inherit the allele are at high risk (Sherman *et al.*, 1984) The disease derives its name from the

observation that a variable percentage of cells from affected patients cytogenetically show a gap at map position Xq27.3 in metaphase spreads under conditions that alter deoxypyrimidine pools (Krawczun et al., 1985; Sutherland and Hecht, 1985).

The gene defective in FRAXA has since been mapped to this same region and subsequently this region was cloned and found to contain an unstable $(CGG)_N$ repeat that was expanded in affected and males and carrier females (Dietrich et al., 1991; Fu et al., 1991; Kremer et al., 1991; Oberlé et al., 1991; Verkerk et al., 1991). This repeat was shown to be polymorphic in the normal population with a range of 6 to 54 repeats while affected individuals show greater than 200 repeats. Interestingly, repeat sizes from approximately 50 to 200 do not result in the Fragile X phenotype yet have a mutation rate close to one and are thus at high risk to expand and pass on the disorder. This range of repeats is referred to as a premutation and it explains the observation of the normal-transmitting male. Dietrich et al. (1991) also showed that a CpG island 250 bp distal to the $(CGG)_N$ repeat was methylated on chromosomes which contained an expanded allele. It was then demonstrated that the $(CGG)_N$ repeat in the Fragile X region was contained within a transcribed sequence and subsequent cloning of the cDNA (FMR-1) has shown that the $(CGG)_N$ repeat is located in the first exon of the FMR-1 gene and methylation of the upstream CpG island leads to a lack of expression of FMR-1 (Oberlé et al., 1991; Pieretti et al., 1991). Although initially unclear it is now known that the $(CGG)_N$ repeat is contained within the 5′ untranslated region of the FMR-1

transcript (Ashley *et al.*, 1993a). Interestingly, analysis of discordant monozygotic twins has shown that the expansion of the repeat appears to occur postzygotically (Devys *et al.*, 1992; Kruyer *et al.*, 1994). Wöhrle *et al.* (1993) offered further support of this observation by demonstrating that the repeat size in clonal cell lines from FRAXA patients are mitotically stable indicating that the mitotic mosaicism of patients must be generated early in development. Additionally, Reyniers *et al.* (1993) have shown that FRAXA males with a full mutation in lymphocytes only have a premutation in sperm samples.

Sequence analysis of the FMR-1 cDNA and putative reading frame has revealed that the protein contains three separate consensus RNA binding domains, an RGG box and two KH boxes (Ashley *et al.*, 1993b; Siomi *et al.*, 1993). Functional analysis of *in vitro* translated protein has revealed that the protein does bind RNA in a sequence specific manner (Ashley *et al.*, 1993b; Siomi *et al.*, 1993) and binds specifically to its own transcript and a subset of mRNA's generated from a human brain cDNA library (Ashley *et al.*, 1993b). Additionally, Ashley *et al.* (1993b) conducted stoichiometry experiments and showed that a single FMR-1 protein binds two RNA molecules. Further support of FMR-1 acting as a RNA binding protein is given by the observation that a previously described point mutation in the FMR-1 gene that resulted in a severe mental retardation (DeBoulle *et al.*, 1993) maps to a conserved isoleucine in the second KH domain of the FMR-1 protein (Siomi *et al.*, 1993).

Analysis of the expression pattern of FMR-1 has demonstrated that the gene is expressed at its highest levels in brain and testes with lower amounts present in a variety of tissues including, heart, lung, placenta, liver, and kidney (Hinds *et al.*, 1993). Additionally, Hinds *et al.* showed by *in situ* hybridization that FMR-1 is highly expressed early in embryonic development and decreases in later stages while becoming more tissue restricted. Further analysis on 25 week human fetal brain showed universal expression of FMR-1 with the nucleus basalis magnocellularis and the hippocampus showing the highest levels (Abitbol *et al.*, 1993). Abitbol *et al.* (1993) also demonstrated that in all regions of brain examined the labeling appeared to be specific to neural cells. Further work by Ashley *et al.* (1993a) has demonstrated that in mouse and human the FMR-1 gene utilizes alternative splicing to generate 12 different transcripts. Six of these transcripts are missing exon 14 which results in a one base pair frameshift that would generate a novel C-terminus. However, Western blot analysis by Siomi *et al.* (1993) has detected only one isoform of the protein which was approximately 80 kDa, a size that is larger than any of the predicted isoforms. It may be that the protein migrates anomolously in SDS-PAGE gels.

Mechanistically, it now appears that a predisposing chromosome that carries an old unstable haplotype is able to expand to a premutation (Richards *et al.*, 1992; Oudet *et al.*, 1993; Smits *et al.*, 1993). Later, upon expansion to a full mutation the upstream CpG island becomes methylated and down regulates the expression of FMR-1. The elimination of expression

then results in mental retardation. This hypothesis is supported by the observation of patients with deletions of the FMR-1 gene (Wöhrle *et al.*, 1992; Meijer *et al.*, 1994) and a single patient with a point mutation (DeBoulle *et al.*, 1993) showing typical mental retardation of the syndrome. Also patients have been documented who have the full expansion yet by isoschizomeric analysis show only partial methylation of the upstream CpG island and are consequently phenotypically normal (McConkie-Russel *et al.*, 1993; Kruyer *et al.*, 1994; Rousseau *et al.*, 1994). Presently, the molecular pathology of FRAXA is the best understood of all of the diseases of trinucleotide repeat expansion and a complete understanding of the function of the FMR-1 protein will eventually lead to deciphering the connection between the molecular defect and the phenotype of mental retardation.

3. **Fragile X Syndrome E (FRAXE)**

Fragile X syndrome E is also characterized by mental retardation although it appears to be milder in form compared to FRAXA (Knight *et al.*, 1993). It was originally described as a separate fragile site that was telomeric to the FRAXA site at Xq28 (Sutherland and Baker, 1992). Additionally, a number of patients were described who showed a FRAXA phenotype yet these patients did not demonstrate an expanded $(CGG)_N$ repeat in the FMR-1 gene (Nakahori *et al.*, 1991; Sutherland and Baker, 1992; Flynn *et al.*, 1993). This work culminated in the cloning of a region of DNA

from Xq28 which carried an unstable $(GCC)_N$ that was expanded in FRAXE patients and carriers (Knight *et al.*, 1993). The $(GCC)_N$ repeat in FRAXE was shown to be polymorphic in the normal population with a range of 6 to 25 repeats (Knight *et al.*, 1993). By Southern blot analysis FRAXE affected males show increases in fragment size of 650 to 2200 bp corresponding to repeat sizes of 200 to over 700 while carrier females showed expansion in the range of 100 to 150 repeats (Knight *et al.*, 1993). Additionally, Knight et al. (1993) showed that a CpG island immediately proximal to the unstable $(GCC)_N$ repeat is methylated in FRAXE affected males. The molecular defect in FRAXE appears to be very similar to FRAXA yet to date no cDNA from the region has been published raising the question is the $(GCC)_N$ repeat in FRAXE located within a transcriptional unit? It would seem likely that the $(GCC)_N$ repeat is located within a transcribed gene as the presence of the nearby CpG island would suggest a possible promoter region. Also, it would seem likely that the $(GCC)_N$ repeat is present in an untranslated region of this yet to be described gene given the size of the expansions observed in FRAXE. Confirmation of this speculation will have to await the cloning of a cDNA from this region.

CHAPTER III

MATERIALS AND METHODS

A. Materials

Enzymes and chemicals were purchased from Ambion (Austin, TX), Amersham (Arlington Heights, IL), BRL (Gaithersburg, MD), Epicentre (Madison, WI), New England BioLabs (Beverly, MA), Pharmacia (Piscataway, NJ), Promega (Madison, WI), Sigma (St. Louis, MO), and Stratagene (La Jolla, CA). Radioisotopes $[\alpha\text{-}^{32}P]dCTP$ (3000 Ci/mmol) and $[\alpha\text{-}^{35}S]dATP$ (3000 Ci/mmol) were purchased from Amersham. Oligonucleotides were obtained from the Wells Center Oligonucleotide Facility (Riley Hospital, Indiana University Medical Center, Indianapolis,IN). Nitrocellulose membranes were purchased from Schleicher and Schuell (Keene, NH). Multiple Tissue Northern Blots were from Clontech (Palo Alto, CA) and the human/rodent somatic cell hybrid panel was from Oncor (Gaithersburg, MD).

B. RNA Extraction

Total cellular RNA was isolated using a modification of a procedure previously described (Chomcyznski and Sacchi, 1987). Samples were homogenized in 500 µl of 4 M guanidinium thiocyanate

(GIT) buffer (4 M guanidine isothiocyanate, 25 mM sodium citrate, pH 7.0, 0.5% sarkosyl and 0.1 M 2-mercaptoethanol). The homogenate was extracted by the addition of the following, 50 µl 2 M sodium acetate (pH 4.0), 500 µl phenol, and 100 µl chloroform:isoamyl alcohol (49:1), mixed, and incubated on ice for 15 minutes. The sample was centrifuged at 10,000 x g for 15 minutes at 4°C. The aqueous phase was transferred to a new tube and the nucleic acid precipitated by the addition of an equal volume of isopropanol and incubated on dry ice for 30 minutes. The RNA was pelleted by centrifugation at 10,000 x g for 15 minutes at 4°C. The isopropanol was aspirated and the resulting pellet resuspended in 100 µl (dependent on size of pellet) of GIT. The pellet was fully dissolved by heating at 65°C and occasionally mixing. RNA was reprecipitated by addition of 0.1 volume 3 M sodium acetate (pH 5.2) and an equal volume of isopropanol, followed by incubation on dry ice for 15 minutes. The RNA was pelleted by centrifugation at 10,000 x g for 15 minutes. This pellet was washed with 500 µl 70% ethanol, dried at 65°C for 2 minutes, and resuspended in 200 µl of diethylpyrocarbonate (DEPC, 0.2%) treated water. Again the pellet was dissolved by heating at 65°C. The final RNA sample was stored at -80°C until needed.

C.   3' RACE

3' RACE was carried out as illustrated in Figure 2. Two separate experiments were carried out using oligos that contained 4 or 8 repeats of CAG, respectively. Reverse transcription was carried out by annealing 3.0 µg of total RNA to 500 ng of
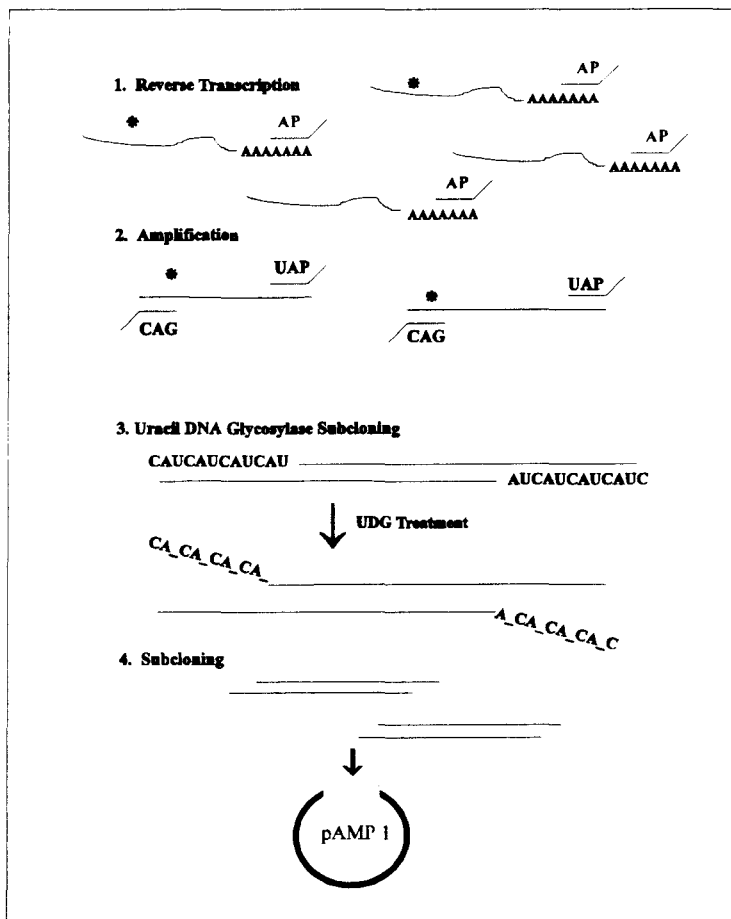
Figure 2: 3' RACE Methodology. The diagram illustrates the methodology utilized for the 3' RACE technique. In step 1 total RNA is reverse transcribed with the adapter primer (AP). An aliquot of the reverse transcription reaction is then used as template in a thermal amplification reaction utilizing the universal amplification primer (UAP) and the CAG primer. The UAP contains sequence that is identical to the 5' portion of the AP primer and allows for amplification from the 3' end of an mRNA. In the work described here two different 3' RACE protocols were carried out utilizing a CAG primer that contained 4 or 8 repeats of CAG. In step 3 the reaction products are treated with Uracil DNA glycosylase to create 12 bp sticky ends and the products are subsequently cloned into the vector pAMP1 (step 4).

oligonucleotide of the sequence [5'-GGC CAC GCG TCG ACT AGT AC(T)$_{16}$-3']. This reaction was incubated for 1 hour at 42°C in the presence of 10 mM Tris pH 8.0, 0.5 mM deoxynucleotide triphosphates, 20 mM dithiothrietol, and 10 units Superscript II reverse transcriptase (BRL). Following reverse transcription, 5.0 µl of the RT reaction was added to a thermal amplification reaction utilizing 10 pmol of each of the primers [5'-(CAU)$_4$(CAG)$_N$-3'](N = 4 or 8) and [5'-(CUA)$_4$GGC CAC GCG TCG ACT AGT AC-3'] in the presence of 50 mM Tris-pH 8.0, 20 mM NH$_4$SO$_4$, 1.0 mM MgCl$_2$, 0.1 mM dNTP's, and 1.0 unit of *Tfl* thermostable polymerase (Epicentre). The 5' oligo contained the CAG repeats and amplified from this sequence within the cDNA population and the 3' oligo was a nested primer complimentary to the oligo used for reverse transcription. Forty cycles of amplification were carried out with a 30 second denaturation at 95°C, a 1 minute annealing at 65°C, and a 2 minute extension at 72°C. The reaction was completed by a final extension at 72°C for 10 minutes. An aliquot of the reaction products were analyzed by agarose gel electrophoresis and visualized by ethidium bromide staining. The remaining portion of the reaction products were subjected to Glassmax purification (BRL) and subsequently batch subcloned using the Uracil DNA Glycosylase (UDG) cloning method (BRL).

D.   Uracil DNA Glycosylase Subcloning

Uracil DNA glycosylase (UDG) cloning was carried out by mixing 50-100 ng of the PCR product, 25 ng of the pAMP 1 vector DNA (25 ng/µl), and 1 U of UDG in a final volume of 20 µl. This

reaction was incubated at 37°C for 30 minutes. The entire UDG reaction was used for transformation.

E.   Transformation of Competent Bacteria Cells

Transformation was carried out utilizing commercially available competent cells as specified by the manufacturers directions (HB101, BRL; JM109, Promega). Briefly, nucleic acids were mixed with 50 μl of competent cells and incubated for 1 hour on ice. The transformation was heat shocked for 20 seconds at 37°C followed by chilling on ice for 2 minutes. The transformation was then incubated for 1 hour in an environmental shaker at 37°C followed by plating on an LB-agar plate containing the appropriate antibiotic. Transformants were analyzed for the presence of recombinant plasmid by PCR or restriction digestion.

F.   Preparation of Frozen Sterile Bacterial Cultures

A single colony of cells was aseptically transferred to a tube containing 2.0 ml of LB medium supplemented with the appropriate antibiotic. This culture was grown overnight at 37°C in an environmental shaker. Cells (800 μl) were placed in a sterile tube and mixed with 200 μl of sterile glycerol. The cells were frozen at -70°C and stored indefinitely. Bacteria were recovered by streaking a sample of the frozen stock onto the appropriate LB agar-antibiotic plate.

G.   Thermal Cycle Amplification of Bacterial Colonies

PCR reactions were carried out with a scrape of a bacterial colony which was heated at 95°C for 10 minutes in 1X reaction buffer(50 mM Tris pH 8.0, 20 mM NH_4SO_4, 1.0 mM $MgCl_2$) to lyse the cells.   10 pmol of each oligonucleotide, 200 µM of each deoxynucleotide triphosphate were added in a final volume of 99 µl. Reactions were brought to 72°C and 1.0 U of *Tfl* DNA polymerase (Epicentre)was added. The reaction mixture was subjected to 25 cycles of 95°C for 30 seconds; 55°C for 1 minute; 72°C for 2 minutes, and a final 72°C elongation period for 10 minutes.   Five microliters of the reaction was analyzed by fractionation in a 1.0% agarose gel containing 0.5 µg/ml ethidium bromide and 1X TBE buffer (20X TBE = 1.78 M Tris-HCl; 1.78 M boric acid; 4 mM EDTA, pH 8.0).   Products were visualized by UV transillumination.

H.   Plasmid DNA Purification

Plasmid DNA was isolated using the alkaline lysis technique as described by Maniatis *et al.*(1989).   A single bacterial colony was inoculated into the appropriate antibiotic containing LB media. The volume of the culture varied based on the amount of plasmid needed.   The culture was grown overnight in an environmental shaker at 37°C.   1.5 ml of overnight culture was transferred to a microcentrifuge tube and the bacteria pelleted by centrifugation at 10,000 x g for 30 seconds.   The media was aspirated off and the bacteria resuspended in 250 µl of ice-cold P1 solution (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 400 µg/ml RNAse A; 250 µl of P1 for each 1.5 ml of bacteria culture).   Once the

pellet was fully resuspended, 250 μl of P2 solution (200 mM NaOH, 1.0% SDS) was added, the tube inverted several times, and the mixture allowed to incubate at room temperature for 5 minutes. Following incubation at room temperature, 250 μl of P3 (2.55 M potassium acetate, pH 4.8) was added and mixed thoroughly. The bacterial cell lysate was centrifuged at 10,000 x g for 15 minutes at 4°C. The resulting supernatant was transferred to a new tube without disrupting the precipitant formed. Columns were utilized to isolate highly purified plasmid DNA (Wizard Plasmid Prep, Promega). Otherwise, DNA was precipitated from the supernatant by the addition of 0.6 volumes of isopropanol. The precipitate was incubated at -20°C for 30 minutes and plasmid DNA pelleted by centrifugation at 10,000 x g for 15 minutes at 4°C. The supernatant was aspirated off and the pellet washed with 70% ethanol. Plasmid was dried at 65°C for 5 minutes and the DNA dissolved in 50 μl of TE (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA). The plasmid DNA solution was placed at 4°C for short-term storage or frozen at -20°C for long-term storage.

I.   **Restriction Digestion and Gel Purification of DNA**
      Fragments

cDNA fragments were isolated by digesting approximately 10 μg of plasmid DNA with 20 U of *Eco*RI and 20 U of *Bam*H1 in appropriate enzyme buffer as described by the manufacturer (BRL) at 37°C for 4 hours. Reactions were terminated by heating at 65°C for 10 minutes. Fragments were separated from vector by agarose gel electrophoresis and excised form the gel. This gel slice was

blotted dry on Whatman 3MM paper and transferred to a punctured sterile 0.7 ml microcentrifuge tube containing 2-3 mm of sterile glass wool. The 0.7 ml tube was placed into a 1.7 ml microcentrifuge tube and centrifuged at 10,000 x g for 10 minutes. The eluant containing the DNA fragment was then analyzed by agarose gel electrophoresis and an approximate concentration determined. Fragments were used in random priming labeling reactions for use in hybridization.

J.   **Preparation of Double Stranded DNA Sequencing Templates**

Double stranded DNA sequencing templates were purified by the method of Majumdar *et al.* (1993). Bacteria colonies were grown overnight in LB containing appropriate antibiotics. 1.0 ml of culture was transferred to a 1.5 ml microcentrifuge and pelleted by centrifugation at 12,000 x g for 30 seconds. Pellets were vortexed for 10 seconds followed by resuspension in 500 μl 1X STET buffer (8% sucrose, 50 mM Tris-HCl (pH 8.0), 50 mM EDTA, and 5% Triton X-100). Lysozyme was added to a final concentration of 0.1 μg/μl and the samples incubated at room temperature for 2 minutes followed by heating at 100°C for 1 minute. Tubes were then centrifuged at 12,000 x g for 10 minutes at 4°C. Pellets were removed using a sterile toothpick and the supernatant was brought up to 500 μl total volume with 1X STET buffer. 10 μl of 10 N NaOH was added to each sample followed by incubation at 37°C for 10 minutes. Following this incubation 400 μl of isopropanol was added and samples incubated 10 minutes at -20°C. Denatured plasmid was pelleted by centrifugation at 12,000 x g for 10 minutes at 4°C.

Pellets were washed with 70% ethanol, dried by heating at 65°C, and resuspended in 16 µl of sterile H_2O. Seven microliters of plasmid was used directly in annealing reactions for DNA sequencing.

K.   **Preparation of Single Stranded DNA Sequencing Templates**

Single stranded DNA sequencing templates were prepared by the method of Russel *et al.* (1986). Bacteria cells (HB101 F[+] or JM109 F[+]) containing pBKS plasmid with fragments of interest cloned into the *EcoRI* site were patched onto LB-Agar plates containing appropriate antibiotics and allowed to grow overnight at 37°C. A portion of each patch was used to inoculate 1.5 ml of 2X YT+G media (10 g yeast extract, 16 g tryptone, and 5 g NaCl per liter plus 0.1% glucose) and R408 helper phage were added (5 X 10^9-10^{10}). Cultures were allowed to grow 4-5 hours with vigorous shaking at 37°C. Following the growth period cultures were transferred to 1.5 ml microcentrifuge tubes and the bacteria pelleted by centrifugation at 12,000 x g for 5 minutes at room temperature. Phage particles were precipitated by mixing the supernatant with 200 µl 2.5 M NaCl, 20% PEG-6000 and incubating 15 minutes at room temperature. Phage were pelleted by centrifugation at 12,000 x g for 10 minutes at room temperature. The supernatant was aspirated and samples were centrifuged and aspirated a second time to remove any remaining PEG-6000. Phage pellets were resuspended in 100 µl 10 mM Tris-HCl (pH 8.0), 2 mM EDTA and extracted with 50 µl Tris buffered phenol. The aqueous phase was mixed with 250 µl of a 25:1 mixture of ethanol:3M sodium acetate and incubated for 15 minutes in a dry ice/ethanol bath. Single stranded DNA was

pelleted by centrifugation at 12,000 X g for 10 minutes at room temperature. The DNA was washed with 70% ethanol , dried by heating at 65°C, and resuspended in 20 μl of TE (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA). Each sample (2.0 μl) was analyzed by agarose gel electrophoresis.

L.   DNA Sequencing

Sequence analysis of purified recombinant plasmid DNA or single stranded DNA was performed using Sequenase 2.0 (USB). of Single stranded DNA (1.0 μg) or double stranded template (5.0 μg) was annealed with 1 pmol of sequencing primer in buffer containing 40 mM Tris-HCl, pH 7.5, 20 mM $MgCl_2$, and 50 mM NaCl by heating for 2 minutes at 70°C followed by a 30 minute incubation at 37°C. A labeling reaction was then carried out by adding unlabeled nucleotides (dCTP, dGTP, dTTP), 10 μCi of α-35S-dATP, and Sequenase 2.0 polymerase to the annealing reaction. Labeling was carried out for 3 minutes at room temperature followed by addition of equal amounts of the labeling reaction to four termination tubes which contained all four dNTP's and one ddNTP, respectively. Termination reactions were carried out for 5 minutes at 37°C. Reactions were stopped by the addition 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol. Reaction products were heated to 75°C and separated on 6% acrylamide, 7.0 M Urea sequencing gels.

Sequencing gels were poured by mixing 50 ml of sequencing gel matrix [6% polyacrylamide (19:1 acrylamide:bis),7.55 M Urea, 1X Sequencing TBE (0.1 M Tris-HCl, 83 mM boric acid, 1 mM EDTA, pH

8.3)] 120 µl of 25% APS and 35 µl of TEMED. This mixture was poured utilizing a 60 ml syringe and the plates clamped. The gel was allowed to polymerize overnight. Electrophoresis of the sequencing reactions was carried out at 40 watts constant power in 1X sequencing TBE. Following electrophoresis sequencing gels were fixed in 10% methanol, 10% acetic acid for 20 minutes, transferred to Whatman 3MM paper and dried at 80°C for 1.0 to 1.5 hours.

M.   Computer Analysis of DNA Sequences

Computer analysis of DNA sequences was carried using several different application programs. Sequences were entered onto the Indiana University Sunflower system utilizing the GCG package. GenBank searches were carried out using the BLAST search program (Altschul *et al.*, 1988). Personal computer analysis was conducted utilizing the software programs DNAsis (Hitachi Software), Generunner (Hastings Software), and Prosis (Hitachi Software).

N.   Labeling Double-Stranded DNA Fragments with Random
Hexamers

Radioactively labeled double stranded DNA molecules, which were used as hybridization probes, were labeled to high specific activity with [α-^{32}P]dCTP (Amersham) utilizing a Decaprime II kit (Ambion). The kit utilizes a modification of the original random priming method (Feinberg and Vogelstein, 1983). Approximately 25 ng of DNA template in 11.5 µl of sterile water was mixed with 2.5 µl of a 10X random decamer solution in a microcentrifuge tube and denatured by heating at 100°C for 3 minutes. The tube was rapidly

chilled on ice for 2 minutes. The labeling reaction was prepared by adding 5 µl of 5X labeling buffer, 5 µl α-^{32}P dCTP (50 µCi, 3000 Ci/mmol), and 1 µl of exo$^-$ Klenow enzyme in a final volume of 25 µl to the denatured template. The labeling reaction was gently mixed and allowed to incubate at 37°C for 10-15 minutes. The reaction was terminated by the addition of 1 µl of 0.5 M EDTA, pH 8.0 and the tube heated to 100°C for 3 minutes. The tube was chilled on ice for 3 minutes and subjected to gel filtration chromatography. Sephadex G-50 solution (40 mg/ml) was autoclaved and brought to a final concentration of 20 mM NaOH, 1 mM EDTA solution. The G-50 spin column was constructed by adding G-50 Sephadex to a 1 ml syringe that had been plugged with sterile glass wool. The Sephadex was packed by centrifugation at 1000 x g for 5 minutes. The denatured, labeled probe was added to the column, spun at 1000 x g for 5 minutes, and collected.

Percent incorporation was determined by comparing the amount of radioactivity left in the column to the radioactivity collected. Specific activity in dpm/µg was calculated by the following computation: (starting label, 50 µCi)x(fraction of label incorporated, 0.5 for 50%)x(2.2 x 10^6 dpm/µCi)x(40, if 25 ng of DNA is being labeled).

O.   Hybridizations

Blots were prehybridized for at least 1 hour at 42°C with 50% formamide, 5X Denhardt's solution, 1% sodium dodecyl sulfate (SDS), 1 M NaCl, 10 mM NaPO$_4$, pH 6.5, 0.1% pyrophosphate, and 250 µg/ml salmon sperm DNA (heat denatured by boiling 10 minutes).

Radioactively labeled cDNA was added to the prehybridization solution and allowed to hybridize 16 to 24 hours at 42°C. Blots were washed three times for 20 minutes each in 0.1X SSC, 0.5% SDS at 65°C. Following washing the blots were wrapped in Saran wrap and autoradiographed by exposure to Hyperfilm MP (Amersham) using intensifying screens at -80°C.

P. Screening a λgt11 cDNA Library with Radioactively labeled DNA Probes

Approximately 4×10^5 plaques were screened by plating 5×10^4 pfu/150 mm plate. Plating bacteria were prepared by inoculating a single Y1090 bacterial colony into 50 ml of LB medium, supplemented with 0.2% maltose and 10 mM $MgSO_4$ in a 250 ml flask and growing overnight in an environmental shaker at 37°C. Infections were conducted by mixing 100 μl of the plating bacteria (10^8 cells) with 50,000 pfu (as determined by titer experiments) from a Jurkat λgt11 cDNA library (Clontech) and incubating for 20 minutes at 37°C. Quickly, 9 ml (for 150 mm plates, 3 ml for 82 mm plates) of prewarmed (42°C) LB soft agarose (LB media + 0.7% agarose) was added to the infection sample and poured onto prewarmed (42°C) LB agar plate (LB media + 1.5% bactoagar). The plate was swirled while pouring to ensure even spreading of the agarose over the plate. Plates were allowed to cool at room temperature for 5 minutes before being placed at 42°C. The plates were incubated until plaques were just beginning to make contact with one another.

Lifts were carried out by placing nitrocellulose filters smoothly onto the plates. The alignment of the filter was marked by asymmetrically stabbing through the filter into the agar with a needle. Filters were incubated on the plates for 1 minute followed by a 1 minute wash in DNA denaturing solution (1.5 M NaCl, 0.5 M NaOH). The filter was partially dried and transferred to neutralizing solution (1.5 M NaCl, 0.5 M Tris-HCl, pH 8.0) for 3 minutes. The filter was then rinsed in 3X SSC (20X SSC = 3 M NaCl; 0.3 M sodium citrate, pH 7.0) for 3 minutes and subsequently placed on Whatman 3 MM paper to dry. DNA was fixed to the membranes by air drying overnight at room temperature. Hybridization was carried out as described above. Plaques representing positive signals were picked and placed in 500 μl of SM (100 mM NaCl; 8 mM $MgSO_4$-H_2O; 50 mM Tris-HCl, pH 7.5; 0.01% gelatin)/0.3% chloroform and stored at 4°C until subsequent screening experiments were performed. Positive clones were purified by successive plating at lower density.

Q.   Purification of λphage DNA and Isolation of Candidate Inserts

Phage DNA was purified by harvesting phage from a confluent plate. Confluent plates were washed overnight with 10 ml of 1X SM at 4°C. The 10 ml of 1X SM was collected and the plates were rinsed with an additional 2 ml of 1X SM. This solution was then centrifuged for 10 minutes at 4,000 X g at 4°C. The resulting supernatant was then brought up to a volume of 10 ml and extracted with 10 μl of chloroform. Following extraction the sample was

centrifuged for 5 minutes at 2,000 rpm at room temperature. The resulting supernatant was transferred to a new tube and equilibrated to 37°C for 5 minutes followed by digestion with 10 µg/ml RNAse A and 8 U/ml DNAse I. Phage was precipitated by the addition of 2.0 ml of 30% PEG 6000, 3.0 M NaCl followed by a one hour incubation on ice. Phage were pelleted by centrifugation at 10,000 X g for 10 minutes. The phage pellet was resuspended in 1.0 ml of 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, and 25 mM EDTA. An equal volume of 4% SDS was added and the mixture incubated 10 minutes at 70°C. 1.0 mL of 2.55 M potassium acetate (pH 4.8) was added and the sample was centrifuged at 15,000 X g for 30 minutes at 4°C. The supernatant was then passed over a Wizard Lambda Prep column (Promega) and washed with 4.0 ml of 80% isopropanol. The λDNA was eluted with 100 µl of 80°C sterile H_2O.

Purified λDNA was digested with *Eco*RI for four hours at 37°C, and the inserts gel purified. These DNA fragments were subsequently ligated into linearized pBKS (Stratagene) that had been digested with *Eco*RI and treated with calf intestinal alkaline phosphatase (CIAP; BRL) as per the manufacturers instructions. The CIAP was inactivated by incubating the reaction at 65°C for 15 minutes. Digested/CIAP treated plasmid was gel purified and used in ligation reactions. Ligations were carried out using a 10:1 ratio of insert:vector in the presence of 1X ligase buffer (10X ligase buffer = 300 mM Tris-HCl, pH 7.5; 100 $MgCl_2$; 100 mM DTT), ATP (final concentration of 0.5 mM), and 1 U of T4 DNA ligase (Promega) in a final volume no greater than 15 µl. Ligations were performed at 14°C overnight. T4 DNA ligase was inactivated by

heating the reaction at 65°C for 10 minutes. The tube was cooled on ice for 2 minutes and the entire reaction transformed into competent bacteria cells.

R.   Chromosomal Mapping By Hybridization to a Human/Rodent Somatic Cell Hybrid Panel

The CAG8-16 (ssbp) full length cDNA fragment was released from pBKS by restriction digestion with *Eco*RI and gel purified. The purified fragment was radioactively labeled as described above and hybridized to a human/rodent somatic cell hybrid panel (Oncor) as described under hybridization. The blot was exposed for four days and the chromosomal assignment made on the basis of the position of the hybridization signal.

S.   Random RACE

The methodology used for the Random RACE technique is summarized in Figure 3. 3.0 µg of total Jurkat RNA was annealed to 500 ng of oligonucleotide of the sequence [5'-GGC CAC GCG TCG ACT AGT AC(X)$_{10}$-3'], where the X indicates a random base. Reverse transcription was carried out for 1 hour at 42°C in the presence of 10 mM Tris pH 8.0, 0.5 mM deoxynucleotide triphosphates, 20 mM dithiothrietol, and 10 units Superscript II reverse transcriptase (BRL). Following reverse transcription, 5.0 µl of the RT reaction was added to a thermal amplification reaction utilizing 10 pmol of each of the primers [5'-(CAU)$_4$(CAG)$_8$-3'] and [5'-(CUA)$_4$GGC CAC GCG TCG ACT AGT AC-3'] in the presence of 50 mM Tris-pH 8.0, 20 mM NH$_4$SO4, 0.1 mM dNTP's, and 1.0 unit of *Tfl* thermostable polymerase

(Epicentre). Forty cycles of amplification were carried out with a 30 second denaturation at 95°C, a 1 minute annealing at 65°C, and a 2 minute extension at 72°C. The reaction was completed by a final 10 minute extension at 72°C. An aliquot of the reaction was analyzed by agarose gel electrophoresis and DNA visualized by ethidium bromide staining. The remainder of the reaction was purified over a Wizard PCR prep column and batch subcloned by the UDG method as described above.
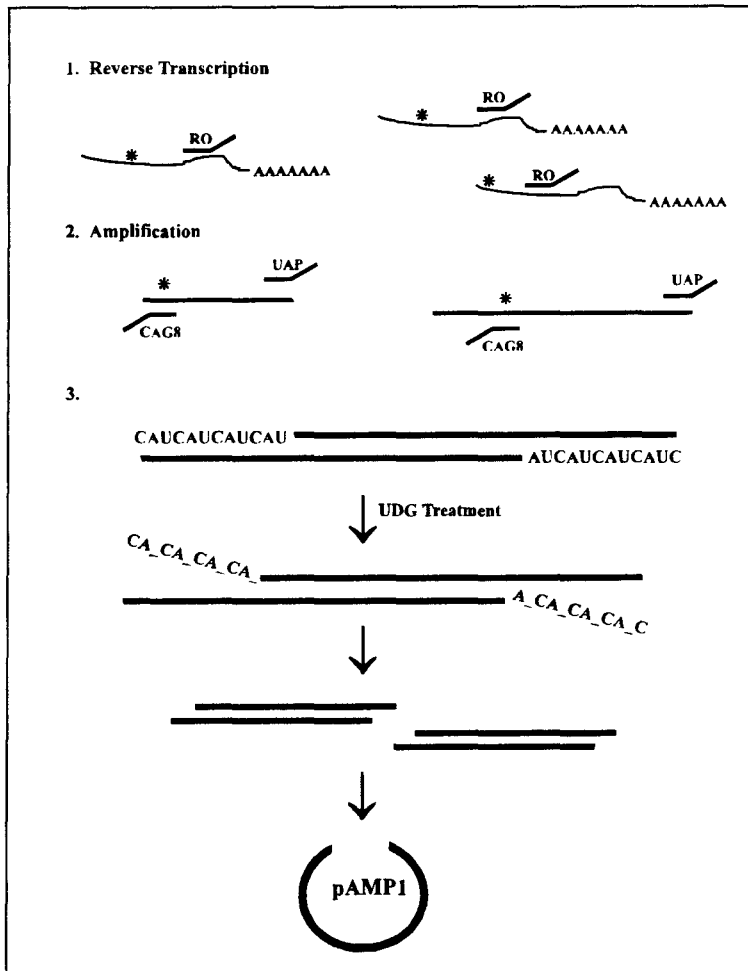
Figure 3: Random RACE Methodology. The diagram illustrates the methodology utilized to carry out the Random RACE experiments. In step 1 total RNA is reverse transcribed with random oligomers (RO) that have a 5' unique sequence region. An aliquot of this reverse transcription is utilized in a thermal amplification utilizing the UAP oligo which is identical to the unique sequence region of the RO and the CAG8 oligo (step 2). In step 3 the reaction products are subjected to Uracil DNA glycosylase treatment (step 3) followed by subcloning into the vector pAMP1. *- Indicates a $(CAG)_N$ repeat within a transcript.

CHAPTER IV

RESULTS

A. 3' RACE Cloning of (CAG)$_N$ Containing cDNA Fragments

1. 3' RACE Cloning with the CAG4 Oligo

Initially I investigated if the 3' RACE technique would be applicable to isolating CAG repeat containing cDNA fragments. Utilizing total RNA from the glioma cell line NAT I was able to isolate five novel clones which contained CAG repeats (Table 3). Additionally, a cDNA fragment was isolated that is identical to the 3' untranslated region of the human $G_{\alpha s}$ cDNA (Mattera *et al.*, 1986). Analysis of this previously isolated cDNA revealed that it contained 3 repeats of CAG in this untranslated region. The isolation of this fragment verified that the technique was operating as intended. These clones will be described individually.

TABLE 3:   CAG4 3' RACE CLONES

Clone	Repeats	Size	GenBank Search
CAG4-3	9	225	No match
CAG4-6	3	93	$G_{\alpha s}$
CAG4-7	7	64	No match
CAG4-10	4	129	No match
CAG4-19	5	78	No match
CAG4-31	6	167	No match

Table 3: CAG4 3' RACE Clones.   The clones isolated by the 3' RACE method utilizing the CAG4 oligo are shown.   All clones were subjected to DNA sequence analysis which revealed the size and the number of CAG repeats.   Five of the clones were found to be novel while clone CAG4-6 is identical to a portion of the 3' UTR of the human $G_{\alpha s}$ cDNA.

a.   Clone CAG4-3

Clone CAG4-3 was cloned by the 3' RACE technique utilizing

the CAG4 oligo. Sequence analysis of the entire clone revealed a

225 bp insert that contained 9 CAG repeats.   A BLAST search

revealed no significant similarity to any sequence in Genbank. The

results of expression analysis of clone CAG4-3 are shown in Figure

4. Using the CAG4-3 insert as a probe on a Human Multiple Tissue

Northern Blot (Clontech) revealed that the gene for CAG4-3

generates a 1.35 transcript that is restricted in its expression,

being detectable only in heart, lung, liver, and kidney. The gene

may be expressed at low levels in brain, placenta, skeletal

muscle, and pancreas but not detectable by the methodology

utilized. Analysis of the three possible reading frames of CAG4-3

reveals no partial peptides with any significant similarity, aside

from the polyglutamine generated by the CAG repeat.

b.   Clone CAG4-6

Clone CAG4-6 was found to contain a 93 bp insert by DNA

sequence analysis.   This insert was shown by   BLAST search

(Altchsul et al., 1990) to be identical to the 3' end of the human

$G_{\alpha s}$ cDNA (Figure 5; Mattera et al., 1986).   This region of the $G_{\alpha s}$
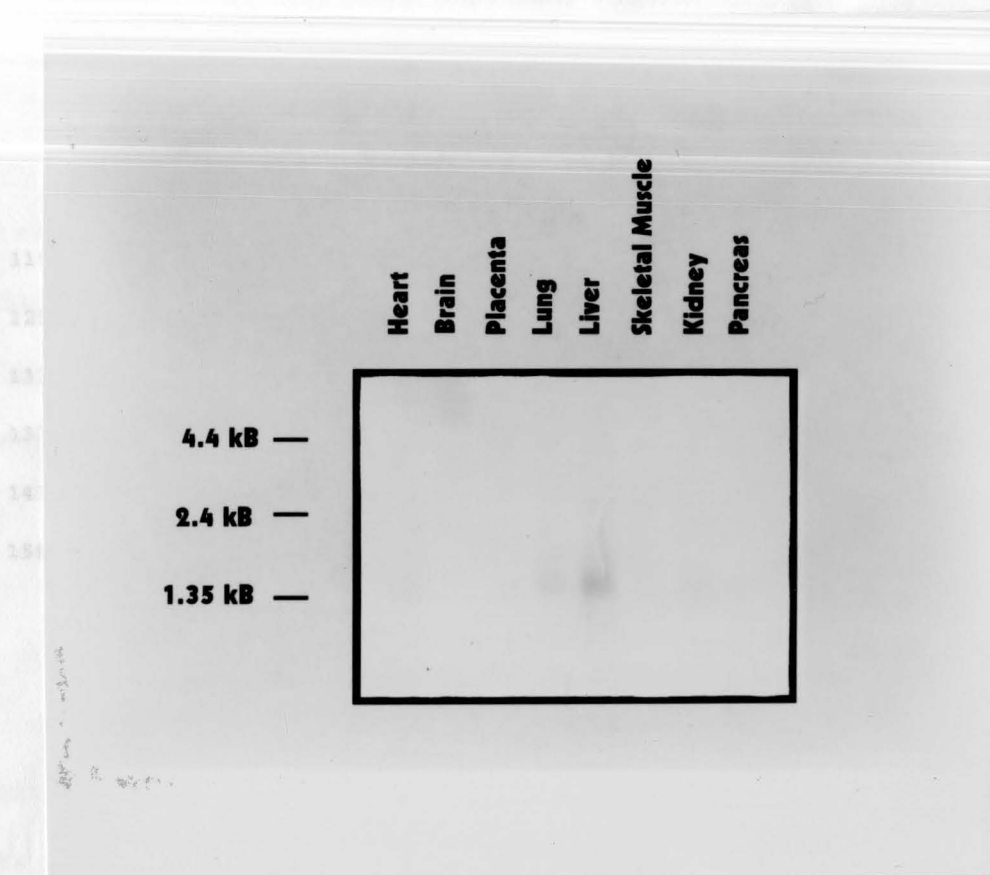
Figure 4: CAG4-3 on Human Multiple Tissue Northern Blot. The result of probing a Human Multiple Tissue Northern Blot with the clone CAG4-3 is shown. The ~1.5 kb transcript appears to be produced in detectable amounts in heart, lung, liver, and kidney.

```
1190   GGAACCCCCA AATTTAATTA AAGCCTTAAG CACAATTAAT TAAAAGTGAA ACGTAATTGT

1251   ACAAGCAGTT AATCACCCAC CATAGGGCAT GATTAACAAA GCAACCTTTC CCTTCCCCCG

1311   AGTGATTTTG CGAAACCCCC TTTTCCCTTC AGCTTGCTTA GATGTTCCAA ATTTAGAAAG

1371   CTTAAGGCGG CCTACAGAAA AAGGAAAAAA GGCCACAAAA GTTCCCTCTC ACTTTCAGTA

1431   AAAATAAATA AAACAGCAGC AGCAAACAAA TAAAATGAAA TAAAAGAAAC AAATGAAATA

1591   AATATTGTGT TGTGCAGCAT TAAAAAAAAT CAAAATAAAA ATTAAATGTG AGCAAAG
```

Figure 5: 3' region of the G$_{\alpha S}$ cDNA. The last 356 bp of the 3' untranslated region of the GαS cDNA are shown. The bases in bold indicate the 93 bp region corresponding to clone CAG4-6 beginning with the three repeats of CAG.

cDNA contains 3 CAG repeats and the cloning of this fragment illustrated that the 3' RACE technique was operating as intended.

c.   Clone CAG4-7

The insert for clone CAG4-7 was found by sequence analysis to contain 64 bp which showed no significant homology to any entries in GenBank.   The clone contained 7 repeats of CAG and given its small size it is suspected that this clone is located in the 3' untranslated region of its respective transcript.

d.   Clone CAG4-10

Clone CAG4-10 contained a 129 bp insert including 4 CAG repeats.   BLAST search (Altschul *et al.*, 1990) revealed no known genes with any significant similarity to this sequence.   This clone was not subjected to further analysis.

e.   Clone CAG4-19

Clone CAG4-19 was found contain 5 repeats of CAG within a 78 bp insert.   BLAST search (Altchul et al., 1990) revealed no known genes with any similarity to CAG4-19.

f.   Clone CAG4-31

CAG4-31 was found by sequence analysis to be a 167 bp insert that contained 6 CAG repeats. BLAST search (Altschul et al., 1990) showed no genes of any significant similarity to CAG4-31.

2.   3' RACE Cloning with the CAG8 oligo

In order to isolate fragments with a larger number of repeats and in potential reading frames we next carried out 3' RACE utilizing an oligo with 8 repeats of CAG to clone cDNA fragments from NAT total RNA. Using this strategy we isolated five novel clones (Table 4A). In general these clones were of larger size and contained a larger number of repeats.

a.   Clone CAG8-1

Clone CAG8-1 was isolated two separate times and by sequence analysis one isolate contained 8 repeats and the other contained 12 repeats. This is a puzzling finding that may indicate that the repeat region of the corresponding gene is polymorphic. A BLAST search (Altschul et al., 1990) showed that CAG8-1 corresponded to an expressed sequence tag that was cloned from an HL-60 cell line.

TABLE 4: CAG8 3′ RACE CLONES

A. CAG8 Clones from Glioma Cell Line NAT

Clone	Repeats	Length	GenBank Search
CAG8-1	8/12[a]	433	EST HL60
CAG8-5	25	379	SRP 14
CAG8-6	6	1,185[b]	EST LG11
CAG8-16	5	816[c]	ssbp
CAG8-27	7	327	EST HHEA56S
CAG8-31	8	464	No match

B. Human Brain CAG8 Clones

Clone	Repeats	Length	GenBank Search
hbCAG8-11	8	239	No match
hbCAG8-14	8	221	No match
hbCAG8-54	8	627	CALM 1

Table 4: CAG8 3′ RACE Clones. The clones isolated by the 3′ RACE experiment with the CAG8 oligo are shown. 4A lists the clones isolated from glioma cell line NAT total RNA. 4B lists clones isolated from human brain (temporal cortex) total RNA. The lengths and number of repeats were characterized by DNA sequence analysis and GenBank searches were carried out utilizing the BLAST program (Altschul *et al.*, 1990). [a]Indicates the number of repeats observed in two separate isolates of CAG8-1. [b]Length indicates the size of a full length cDNA isolated from a retinal epithelium cDNA library (Gieser *et al.*, 1992). [c]Length indicates the size of a cDNA isolated from a Jurkat cDNA library.
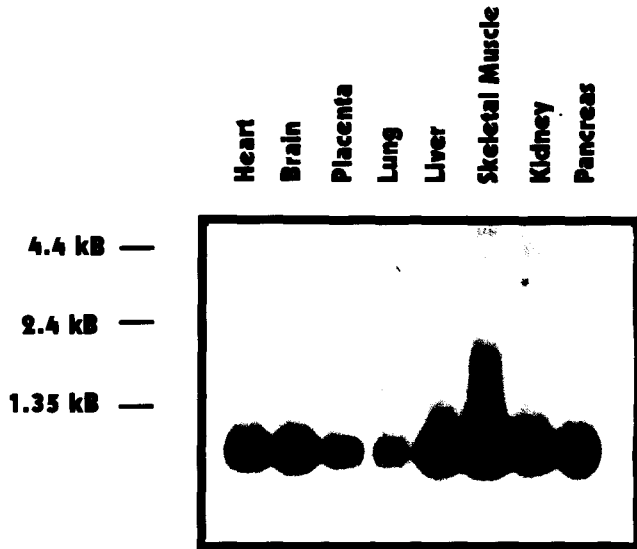
Figure 6: CAG8-1 on Human Multiple Tissue Northern Blot. The result of probing a Human Multiple Tissue Northern Blot (Clontech) with clone CAG8-1 is shown. The ~800 bp transcript is expressed in all tissues examined.

Analysis of the reading frames generated from this sequence does not generate a peptide with any known motifs.

The results of probing a Multiple Tissue Northern blot (Clontech) with the CAG8-1 cDNA fragment are shown in Figure 6. The gene corresponding to this cDNA fragment generates an approximately 800 bp transcript that expressed in all the tissues examined. This data further supports the technical feasibility of 3′ RACE as a methodology for cloning trinucleotide repeat containing cDNA fragments.

b.   Clone CAG8-5

Clone CAG8-5 was also isolated several times, however, there was always a consistent number of repeats that alternated between CAG and CAA (Figure 7). A search of Genbank revealed that this cDNA was recently isolated and is known as signal recognition particle subunit 14 (SRP14). The entire sequence of the SRP14 cDNA with the predicted reading frame is shown in Figure 7. The region corresponding to CAG8-5 is shown in bold and is made up of the 3′ 381 bp. The repeat region is apparently translated as alternating dipeptides of alanine and threonine with one proline interruption (Figure 7).

The result of probing a Human Multiple Tissue Northern Blot (Clontech) with the CAG8-5 insert is shown in Figure 8. SRP14 is

```
1                      M   V   L   L   E   S   E   Q   F   L   T   E   L   T              .

1     TCGAGCCAGCGTCGCCGCGATGGTGTTGTTGGAGAGCGAGCAGTTCCTGACGGAGCTGAC

15       R   L   F   Q   K   C   R   T   S   G   S   V   Y   I   T   L   K   K   Y   D
61    CAGACTTTTCCAGAAGTGCCGGACGTCGGGCAGCGTCTATATCACCTTGAAGAAGTATGA

35       G   R   T   K   P   I   P   K   K   G   T   V   E   G   F   E   P   A   D   N
121   CGGTCGAACCAAACCCATTCCAAAGAAAGGTACTGTGGAGGGCTTTGAGCCCGCAGACAA

55       K   C   L   L   R   A   T   D   G   K   K   K   I   S   T   V   V   S   S   K
181   CAAGTGTCTGTTAAGAGCTACCGATGGGAAGAAGAAGATCAGCACTGTGGTGAGCTCCAA

75       E   V   N   K   F   Q   M   A   Y   S   N   L   L   R   A   N   M   D   G   L
241   GGAAGTGAATAAGTTTCAGATGGCTTATTCAAACCTCCTTAGAGCTAACATGGATGGGTT

95       K   K   R   D   K   K   N   K   T   K   K   T   K   A   A   A   A   A   A   A
301   GAAGAAGAGAGACAAAAAGAACAAAACTAAGAAGACCAAAGCAGCAGCAGCAGCAGCAGC

115      A   A   P   A   A   A   A   T   A   A   T   T   A   A   T   T   A   A   T   A
361   AGCAGCACCTGCCGCAGCAGCAACAGCAGCAACAACAGCAGCAACAACAGCAGCAACAGC

135      A   Q
421   AGCACAGTAAAGGGCATACATTTCCTGCTTTCACCAATTAACCACTGAATTGCTATTTTT

481   TCCTTTTGGCCAGATAGCTAGGTTTCTGGTTCCCCCACAGTAGGTGTTTTCACATAAGAT

541   TAGGGTCCTTTTGGAAAGAATAGTTGCAGTGTTTATAGGATAGTTGTGGTAAGAATCTAG

601   TTTATTTTGCATTTGGCTAATTGGTCTGTGCTGCATGGTTATATACTCCTGGATTATAGA

661   TTAAAAGTCTCTGTAGACATCTCTGTGAAGAGCAAGCTATCATTAAACATGTCTGTTTAT

721   C
```

Figure 7: Sequence of CAG8-5. The sequence of entire 721 bp
signal recognition particle subunit 14 cDNA is shown. The 380 bp
sequence of clone CAG8-5 is indicated in bold. The region of
(CAG)$_N$ repeats is interrupted by CAA repeat units and the entire
repeat region is translated as alanine (A) or threonine (T) with
one proline (P) interruption.
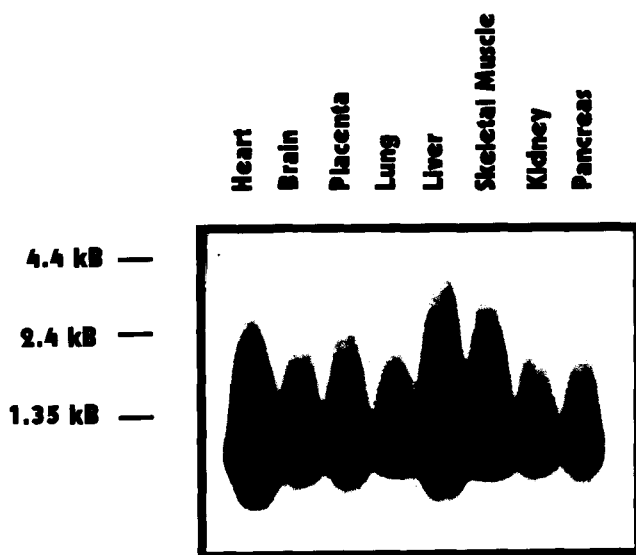
Figure 8: CAG8-5 on Human Multiple Tissue. The result of probing a Human Multiple Tissue Northern Blot with clone CAG8-5 is shown. The ~900 bp transcript is highly expressed in all tissues examined.

ubiquitously expressed at a very high level in all of the tissues examined. Presently, the function of the protein is unclear although it may have a role in mitochondrial protein import.

c.   Clone CAG8-6

Clone CAG8-6 was isolated once with the 8 repeats from the oligo present. Sequence analysis followed by BLAST search (Altschul et al., 1990) revealed that the fragment was identical to the expressed sequence tag LG11 that had been isolated from a subtracted retinal epithelium cDNA library (Geiser *et al.*, 1992). This cDNA clone was obtained from Dr. Anand Swaroop (University of Michigan) and sequenced. The sequence of the 1,185 bp insert with its 819 bp open reading frame that codes for a predicted 273 amino acid protein is shown in Figure 9. The predicted 31,416 Da protein has no significant similarity to any known proteins nor does it contain any obvious motifs. The CAG repeat is made up of 5 CAG units interrupted by a single CAA. This region is translated as serine and arginine (Figure 9). Interestingly, the CAG8-6 protein is made up of over 10% of both of these amino acids. The functional significance of this observation is presently unknown, however, this arginine and serine content is well above the average for known proteins (Doolittle, 1986).

The result of probing a Human Multiple Tissue Northern Blot (Clontech) with the full length CAG8-6 cDNA are shown in Figure

```
1        AAGCCGGCGGCAGCTGCTTGGGGCGGTGCGGTGGTGACTGAGCTACGAGCCTGGCGCGGT

61       GTGCGCCGAGCCCCGGCCGGCCCGCCCTCGCGTGCCTCCCAGCTCCGCACCCCTGATGCT

                                             M   V   K   Y   F   L   G   Q   S
121      GCGCGGGTGCTGAGCCCGCTTCGGCCGGGACGATGGTGAAGTATTTCCTGGGCCAGAGCG

10       V   L   R   S   S   W   D   Q   V   L   P   P   S   F   W   Q   R   Y   P   N
181      TGCTCCGGAGTTCCTGGGACCAAGTGTTGCCGCCTTCTTTCTGGCAGCGGTACCCGAATC

30       P   Y   S   K   H   V   L   T   E   D   I   V   H   R   E   V   T   P   D   Q
241      CCTATAGCAAACATGTCTTGACGGAAGACATAGTACACCGGGAGGTGACCCCTGACCAGA

50       K   L   L   S   R   R   L   L   T   K   T   N   R   M   P   R   W   A   E   R
301      AACTGCTGTCCCGGCGACTCCTGACCAAGACCAACAGGATGCCACGCTGGGCCGAGCGAC

70       L   F   P   A   N   V   A   H   S   V   Y   C   L   E   D   S   I   V   D   P
361      TATTTCCTGCCAATGTTGCTCACTCGGTGTACTGCCTGGAGGACTCTATTGTGGACCCAC

90       Q   N   Q   T   M   T   T   F   T   W   N   I   N   H   A   R   L   M   V   V
421      AGAATCAGACCATGACTACCTTCACCTGGAACATCAACCACGCCCGGCTGATGGTGGTGG

110      E   E   R   C   V   Y   C   V   N   S   D   N   S   G   W   T   E   I   R   R
481      AGGAACGATGTGTTTACTGTGTGAACTCTGACAACAGTGGCTGGACTGAAATCCGCCGGG

130      E   A   W   V   S   S   L   F   G   V   S   R   A   V   Q   E   F   G   L
541      AAGCCTGGGTCTCCTCTAGCTTATTTGGTGTCTCCAGAGCTGTCCAGGAATTTGGTCTTG

150      A   R   F   K   S   N   V   T   K   T   M   K   G   F   E   Y   I   L   A   K
601      CACGGTTCAAAAGCAACGTGACCAAGACTATGAAGGGTTTTGAATATATCTTGGCTAAGC

170      L   Q   A   R   P   L   P   K   H   L   L   R   Q   P   R   K   P   R   R   R
661      TGCAAGCGAGGCCCCTTCCAAAACACTTGTTGAGACAGCCAAGGAAGCCAAGGAGAAGGC

190      Q   R   R   R   H   W   Q   L   Q   R   R   P   R   T   C   P   A   R   R   P
721      AAAGGAGACGGCACTGGCAGCTACAGAGAAGGCCAAGGACCTGCCCAGCAAGGCGGCCAC

210      P   R   S   S   S   S   S   N   S   L   C   S   Q   S   T   T   T   T   A   P
781      CAAGAAG**CAGCAGCAGCAGCAACAG**TTTGTGTAGCCAGTCTACCACCACCACAGCACCCC

230      Q   T   A   S   L   A   P   L   P   F   P   S   L   Y   F   I   I   K   N   Q
841      AGACAGCTAGCTTAGCCCCTCTGCCCTTCCCTTCATTGTACTTTATCATTAAAAATCAAC

250      L   P   A   L   S   A   V   Y   V   V   G   C   G   D   A   V   G   I   C   S
901      TTCCAGCCCTGTCTGCTGTCTACGTGGTGGGTTGTGGGGATGCAGTTGGCATCTGCAGTA

270      T   P   S   T
961      CACCAAGCACATGATTCATGTCTGAGCCAGGTCTGCTTATTCTCCATTGGCAGCTGAGGA
```

1021 CCGAGGCACAGAGGTGCGGTGACTTGCCCGGGCTCCAGGTAGCCTGCAGGTTAACTGGCG

1081 GTAAGTGCTAGACTGTAAGCCCGACAAGGGCAGGGCTTTTGGTTTTGTTCTCTGATGTGT

1141 CTCAGTATCTAGCACATAATAGACACTCAATAAATACTTGTTGAA

Figure 9: Sequence of clone CAG8-6. The entire sequence of the CAG8-6 cDNA is shown. The 1,185 bp cDNA contains an 819 bp open reading frame that codes for a 273 amino acid protein. GenBank search has revealed no known genes with significant homology to CAG8-6. The $(CAG)_N$ repeat is indicated in bold and is translated as serine (S).
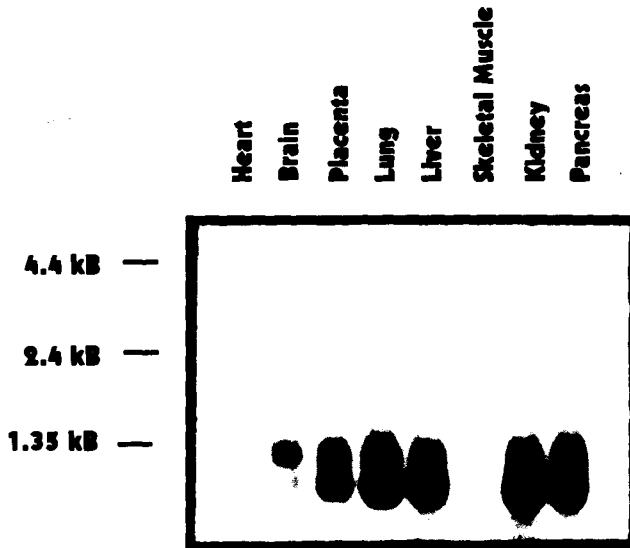
Figure 10: CAG8-6 on a Human Multiple Tissue Northern Blot. The result of probing a Human Multiple Tissue Northern Blot with clone CAG8-6 is shown. This cDNA detects two transcripts of ~1.5 kb and ~1.2 kb. expressed in all tissues examined.

10. The insert detects two transcripts of 1.5 kb and 1.2 kb that are ubiquitously expressed. PCR primers made to the initiation and termination codon regions were used in reverse transcription-polymerase chain reaction and detected only a single 830 bp band (data not shown). This data indicates that if the two transcripts are generated by alternative splicing it is does not occur within the translated region. It does not seem likely that the cDNA is detecting another member of a gene family on Northern blots as the CAG8-6 gene has been mapped by fluorescent *in situ* hybridization to a single locus at chromosome 1q41-42 (A. Swaroop, personal communication). Interestingly, this region has been shown by linkage analysis (Kimberling *et al.*, 1990; Lewis *et al.*, 1990) to be the location of the gene that is defective in type II Usher syndrome. In order to investigate if expansion of the trinucleotide repeat in CAG8-6 was involved in the disease etiology the repeat region was examined in patient DNA by PCR. The region was found to be stable in patients and in the normal population, showing no polymorphism (data not shown). Furthermore, Kimberling and coworkers have recently isolated a candidate gene from the 1q41-42 region that appears to be the gene that is defective in Type II Usher syndrome (personal communication). This data indicates that CAG8-6 is not involved in the molecular pathology of Type II Usher syndrome.

d.   Clone CAG8-16

Clone CAG8-16 was isolated multiple times but never contained
more than the 8 repeats of the CAG8 oligo.   A Genbank search
revealed that this clone is apparently the human homologue to a
mouse cDNA which codes for a single stranded DNA binding protein
(Ballard *et al.*, 1988).   Additionally, a rat homologue was
previously characterized as a cDNA of unknown function which was
highly expressed in tumor cell lines (Soma *et al.*, 1984).   The
CAG8-16 3' RACE clone was used as a probe to screen a Jurkat cell
cDNA library and a single positive clone was isolated.   Sequence
analysis revealed that the 816 bp cDNA contained a 381 bp open
reading frame that coded for a 14,397 Da protein.   The CAG repeat
in CAG8-16 is translated as ACG and therefore codes for serine
(Figure 11).   The high degree of conservation between the human,
rat, and mouse peptides is shown in Figure 11.   The human and
mouse are 93% identical while the human and rat proteins are 94%
identical, not including the 8 N-terminal amino acids that were
not reported in the rat clone (Soma *et al.*, 1984) and probably
indicate an incomplete cDNA.   The work of Ballard *et al.* (1988) on
the biochemical properties of the mouse protein product they
termed p9 indicates that the mouse protein is made as a 15 kDa
precursor that does not bind single stranded DNA that is
proteolytically processed to a 9 kDa precursor which actively

```
                   5           10          15          20          25          30
hssbp    1   M P K S K E L V S S S S S G S D S D S E V D K K L K R K K Q
rssbp    1   X X X X X X X X S S S S S G S D S D S E V E K K L K R K K Q
mssbp    1   M P K S K E L V S S S S S G S D S D S E V E K K L K R K K Q

hssbp   31   V A P E K P V K K Q K T G E T S R A L S S S K Q S S S S R D
rssbp   31   V V P E K P V K K Q K P S E S S R A L A S S K Q S S S S R D
mssbp   31   A V P E K P V K K Q K P G E T S R A L A S S K Q S S S S R D

hssbp   61   D N M F Q I G K M R Y V S V R D F K G K V L I D I R E Y W M
rssbp   61   D N M F Q I G K M R Y V S V R D F K G K I L I D I R E Y W M
mssbp   61   D N M F Q I G K M R Y V S V R D F K G K I L I D I R E Y W M

hssbp   91   D P E G E M K P G R K G I S L N P E Q W S Q L K E Q I S D I
rssbp   91   D S E G E M K P R R K G I S L N M E Q W S Q L K E Q I S D I
mssbp   91   D S E G E M K P G R K G I S L N M E Q W S Q L K E Q I S D I

hssbp  121   D D A V R K L
rssbp  121   D D A V R K L
mssbp  121   D D A V R K L
```

Figure 11: ssbp Sequence Comparison. The human ssbp (CAG8-16) was
compared to both the rat and mouse homologues. The rat sequence
was apparently incomplete and did not contain the amino terminus.
The human protein is 93% identical to the mouse homologue and 94%
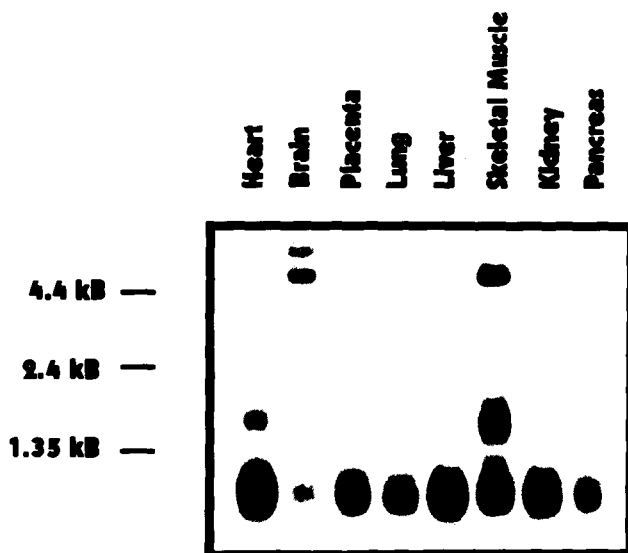identical to the rat homologue.

Figure 12: CAG8-16 Northern Blot. The results of probing a Human Multiple Tissue Northern Blot (Clontech) with the full length CAG8-16 cDNA clone are shown. This cDNA detects three transcripts of 4.4 kb, 1.5 kb, and 0.8 kb expressed in all tissues examined. Additionally, an ~5.0 kb transcript is produced only in brain.

binds single stranded DNA with a greater affinity for RNA. Additionally, the protein appears to have a higher affinity for single stranded DNA that contains 5-methyl cytosine. The cellular function of this activity is presently unclear.

The 816 bp CAG8-16 cDNA was radioactively labeled and used as a probe on a Multiple Tissue Northern Blot (Clontech). The CAG8-16 gene was found to produce three transcripts of 4.4 kb, 1.6 kb, and 0.9 kb(Figure 12). Additionally, a fourth transcript of ~5.0 kb was present only in brain (Figure 12). The cDNA clone isolated from the Jurkat cDNA library appears to correspond to the smallest transcript while the origin of the larger transcripts is unknown. However, Ballard *et al.* (1988) reported a similar finding with the mouse p9 gene and showed that all three transcripts were produced by transcriptional read-through and the use of alternative polyadenylation sites. It seems likely that a similar situation exists for the human gene.

The CAG8-16 816 bp cDNA was radioactively labeled and hybridized to a human/rodent somatic cell hybrid panel (Oncor). The result of this experiment with the hybridization signal indicating that the gene for CAG8-16 resides on chromosome 5 is shown in Figure 13. A search of diseases that have been linked to chromosome 5 has given no obvious phenotypes that could result from a mutation in CAG8-16.

73
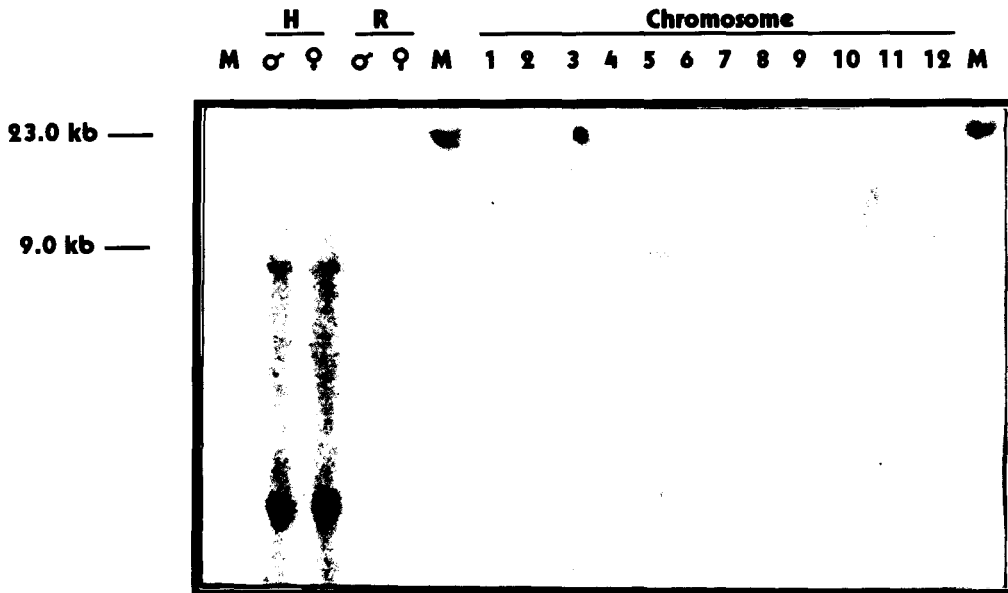
Figure 13: CAG8-16 Chromosomal Localization. The result of probing a human Somatic Cell Hybrid Panel (Oncor) with the full length CAG8-16 cDNA is shown. M - λ/*Hind*III marker, H - Human, R - Rodent. The chromosomes are indicated by the numbers above the respective lanes and the sizes of the marker are indicated on the left. The positive hybridization signal indicates that the gene for CAG8-16 resides on chromosome 5.

```
  1    CAGCAGCAGC AGCAGCAGCA GCAGCCCTGT GGCCGGCACA GAGATCAAGA AACAAAAGCC

 61    AGGCCGAGCA TGGTGGCTCC CGGCCTGTAA TCCCAGCACT TTGGGAGGTC GAGGCAGGCA

121    GTCACTTGAG TTCAGGAGTT CAAGACCAGC CTGGGCAACA TGGCACAACA CATCTCTACA

181    AAAAATACAA AACTTAGCCT CCTGGTGTGG TGGCATGTGC CTGTGGTCCC AGCTGCTTGG

241    GAGGCTGAGA TAGAGATCAC TTGAGCCTGG GAGTTCGAGG CTACAGTGAG CTGTGATTGT

301    ACCACTGCAC TCCACATTGG GCAACAGAAC
```

Figure 14: CAG8-27 Sequence. The sequence of clone CAG8-27 is shown. The sequence highlighted in bold indicates the region that matches with a large number of sequences from GenBank. This region shows similarity to an Alu-like repetitive element.

e.   Clone CAG8-27

Clone CAG8-27 was sequenced and by GenBank search was revealed to show a substantial amount of similarity to a number of sequences. CAG8-27 has a high degree of similarity to a number of GenBank entries due to an Alu-like repetitive element that makes up the majority of the clone (Figure 14). Hybridization to a Human Multiple Tissue Northern Blot was attempted with the clone. However, presumably due to the presence of this repetitive element, strong background hybridization was observed making this result uninterpretable (data not shown).

f.   Clone CAG8-31

Clone CAG8-31 was isolated once by 3' RACE cloning and by sequence analysis was found to contain an insert of 464 bp (Figure 15). Genbank search with this sequence revealed no similarity to any known sequence. Analysis of the open reading frame created by translating the CAG repeat as polyglutamine reveals a protein that would have a C-terminal region that would be very glutamine rich (Figure 15) even without the repeat region. Glutamine rich regions of proteins are known to be present in a number of transcriptional activators and may be important as domains necessary for protein-protein interactions (Gerber *et al.*, 1994).

```
1        Q  Q  Q  Q  Q  Q  Q  Q  L  L  P  S  P  A  A  K  Q  Q  P  A
1     CAGCAGCAGCAGCAGCAGCAGCAGCTCTTGCCATCTCCAGCTGCCAAACAGCAGCCTGCC

21       G  Q  E  Q  P  Q  A  R  E  A  P  G  P  A  Q  G  C  S  C  Q
61    GGGCAGGAGCAGCCCCAGGCCAGAGAGGCTCCCGGTCCAGCTCAGGGATGCTCCTGCCAG

41       H  R  G  Q  G  L  L  E  Q  A  H  S  E  P  G  Q  P  C  Q  S
121   CACAGGGGCCAGGGACTCCTGGAGCAGGCACATAGTGAGCCCGGGCAGCCCTGCCAGTCA

61       G  P  F  P  S  P  L  R  L  G
181   GGCCCCTTTCCTTCCCCATTGAGGTTGGGGTAGGTGGGGGCGGTGAGGGCTCCACGTTGT

241   CAGCGCTCAGGAATGTGCTCCGGCAGAGTGCTGAAGCCTATAACCCCAACCATTTCCTTC

301   GGACGCCCGGTACTCAGCTGGCCCACTCCACAGCCAGCCTGCCCTGCCCTTCACCGTGGA

361   TGTTTTCAGAAGTGGCCATCGAGAGGTCTGGATGGTTTTATAGCAACTTTGCTGTGATTC

401   CCGTTTGTATCTGTAAATATTTGTTCTATAGATAAGATACAAAT
```

Figure 15: CAG8-31 Sequence. The sequence for clone CAG8-31 is shown. The putative reading frame is predicted by translating the CAG repeats as glutamine (Q). This putative reading frame generates a C-terminal peptide very rich in glutamine (Q).
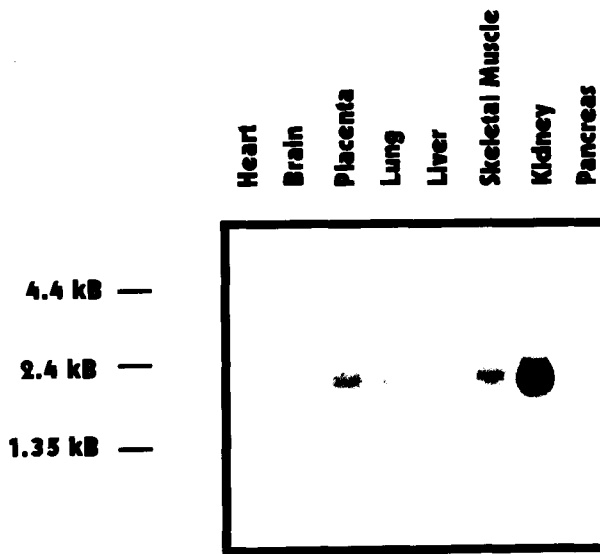
Figure 16: CAG8-31 on Human Multiple Tissue Northern Blot. Clone CAG8-31 detects an ~2.2 kb transcript that is expressed in all tissues examined.

The CAG8-31 clone was radioactively labeled and hybridized to a Human Multiple Tissue Northern blot. The results of this experiment are shown in Figure 16. The CAG8-31 gene generates a 2.2 kb transcript that is ubiquitously expressed and is apparently at its highest level in kidney.

3. 3' RACE Cloning from Human Brain RNA with the CAG8 Oligo

Utilizing an identical approach as described above for the CAG8 cloning, we carried out 3' RACE on total RNA from human temporal cortex. The clones obtained are summarized in Table 4 B and will be discussed individually in the following sections.

a. Clone hbCAG8-11

Clone hbCAG8-11 was isolated and subject to sequence analysis. GenBank search revealed that the 239 bp insert was not similar to any known sequence. Analysis of possible C-terminal peptides showed no regions with significant similarity to any known motifs or proteins.

```
                        5          10         15         20         25
hbCAG8-14      1  Q Q Q Q Q Q Q Q R P [L] H F P L A P [L] L G R E A - L
LZ                - - - - - - - - - - [L] X X X X X X [L] X X X X X X L

                        30         35
hbCAG8-14      26  S G T S L Q [L] P L H L G M [S]
LZ                 X X X X X X [L] X X X X X X [L]
```

Figure 17: Leucine Zipper of Clone hbCAG8-14.   The potential
leucine zipper motif of hbCAG8-14 is shown.   The hbCAG8-14 is
aligned with the consensus leucine zipper which consists of
leucine residues repeated every seven amino acids.   hbCAG8-14 has
an imperfect fit in that one repeat is only six amino acids and
the final expected leucine is serine.   Replacement of a leucine
residue with other amino acids has been observed (Kouzarides and
Ziff, 1988; Landschulz et al., 1989). LZ- leucine zipper.   The
boxes indicate the conserved leucine residues.

b.   Clone hbCAG8-14

Clone hbCAG8-14 was isolated and sequenced and found by GenBank search to be novel. Analysis of possible open reading frames within the 221 bp insert showed that translating the insert with the CAG repeat coding for glutamine results in a 4,240 Da peptide that is over 20% leucine. Analysis of this region indicates that were this the correct reading frame this region could potentially code for a leucine zipper motif (Kouzarides and Ziff, 1988; Landschulz *et al.*, 1989). This potential motif is shown in Figure 17.

The hbCAG8-14 clone was radioactively labeled and hybridized to a Human Multiple Tissue Northern blot (Clontech). This cDNA fragment detects a 1.0 kb transcript that is expressed in all tissues examined (Figure 18). Furthermore, the cDNA appears to weakly hybridize to a number of other larger transcripts in all tissues. The origin of this hybridization is unknown but may indicate that hbCAG8-14 is a member of a gene family. The background hybridization is not thought to arise from the cross-hybridization of the CAG repeat as it was not observed with any other clones.

Figure 18: hbCAG8-14 on Human Multiple Tissue Northern Blot. Clone hbCAG8-14 was used to probe a Human Multiple Tissue Northern Blot. The cDNA detects an ~800 bp transcript that is expressed in all tissues examined. Additionally, the clone detects a number of minor hybridization products. This result may indicate the existence of a gene family.

```
  1 CAGCAGCAGCAGCAGCAGCAGCAGCATCGGAGGTACCCCCGCCGTCGCAGCCCCCGCGCT


  1                                                M   A   D   Q
 61 GGTGCAGCCACCCTCGCTCCCTCTGCTCTTCCTCCCTTCACTCGCACCATGGCTGATCAG

  5 L   T   E   E   Q   I   A   E   F   K   E   A   F   S   L   F   D   K   D   G
121 CTGACCGAAGAACAGATTGCTGAATTCAAGGAAGCCTTCTCCCTATTTGATAAAGATGGC

 25 D   G   T   I   T   T   K   E   L   G   T   V   M   R   S   L   G   Q   N   P
181 GATGGCACCATCACAACAAAGGAACTTGGAACTGTCATGAGGTCACTGGGTCAGAACCCA

 45 T   E   A   E   L   Q   D   M   I   N   E   V   D   A   D   G   N   G   T   I
241 ACAGAAGCTGAATTGCAGGATATGATCAATGAAGTGGATGCTGATGGTAATGGCACCATT

 65 D   F   P   E   F   L   T   M   M   A   R   K   M   K   D   T   D   S   E   E
301 GACTTCCCCGAATTTTTGACTATGATGGCTAGAAAAATGAAAGATACAGATAGTGAAGAA

 85 E   I   R   E   A   F   R   V   F   D   K   D   G   N   G   Y   I   S   A   A
361 GAAATCCGTGAGGCATTCCGAGTCTTTGACAAGGATGGCAATGGTTATATCAGTGCAGCA

105 E   L   R   H   V   M   T   N   L   G   E   K   L   T   D   E   E   V   D   E
421 GAACTACGTCACGTCATGACAAACTTAGGAGAAAAACTAACAGATGAAGAAGTAGATGAA

125 M   I   R   E   A   D   I   D   G   D   G   Q   V   N   Y   E   E   F   V   Q
481 ATGATCAGAGAAGCAGATATTGATGGAGACGGACAAGTCAACTATGAAGAATTCGTACAG

145 M   M   T   A   K
541 ATGATGACTGCAAAATGAAGACCTACTTTCAACTCCTTTTTCCCCCCTCTAGAAGAATCA

601 AATTGAATCTTTTACTTACCTCTTGCAAAAAAAAGAAAAAAGAAAAAAGTTCATTTATTC

661 ATTCTGTTTCTATATAGCAAAACTGAATGTCAAAAGTACCTTCTGTCCACACACACAAAA

721 TCTGCATGTATTGGTTGGTGGTCCTGTCCCCTAAAGATCAAGCTACACATCAGTTTTACA

781 ATATAAATACTTGTACTACCTTAATGATAAGGACTCCTTA
```

Figure 19: The sequence of the human CALM 1 cDNA with predicted reading frame. The sequence initiates with a run of eight CAG repeats (bold) in the 5' untranslated region of the cDNA. In the original *CALM 1* clone (Wawrynczak and Perham, 1984) there was only one CAG reported. The underlined region of poly A+ within the sequence indicates where reverse transcription initiated.

d.   <u>Clone hbCAG8-54</u>

Clone hbCAG8-54 was sequenced and by Genbank search was found to have a sequence identical to the calmodulin 1 (*CALM 1*) cDNA (Wawrynczak and Perham, 1984).   This was initially a puzzling result, considering that the original *CALM I* sequence had no CAG repeats. However, Figure 19 shows that the CAG repeat was at the immediate 5' end of the previously published sequence (Wawrynczak and Perham, 1984).   This led us to conclude that the CAG repeat was present in the 5' untranslated region of the *CALM 1* cDNA but had not been previously described because the gene for *CALM 1* had not been cloned and the transcriptional start site had not been mapped. Furthermore, clone hbCAG8-54 was generated by reverse transcription that initiated at an internal stretch of poly A that is interrupted with a single G (Figure 19).   Rhyner *et al.* (1994) have since cloned the *CALM 1* gene and mapped its transcriptional start site and the results of these two experiments have confirmed that there is a CAG repeat present in the 5' UTR of the *CALM 1* gene.

Clone hbCAG8-54 was used as to probe a Human Multiple Tissue Northern blot. The results of this experiment are shown in Figure 20.   The CALM 1 gene generates three transcripts of 4.0 kb, 1.5 kb, and 0.8 kb. Ni *et al.* (1992) have previously shown that the rat *CALM 1* gene generates two transcripts by the use of alternative polyadenylation sites.   It was hypothesized that an

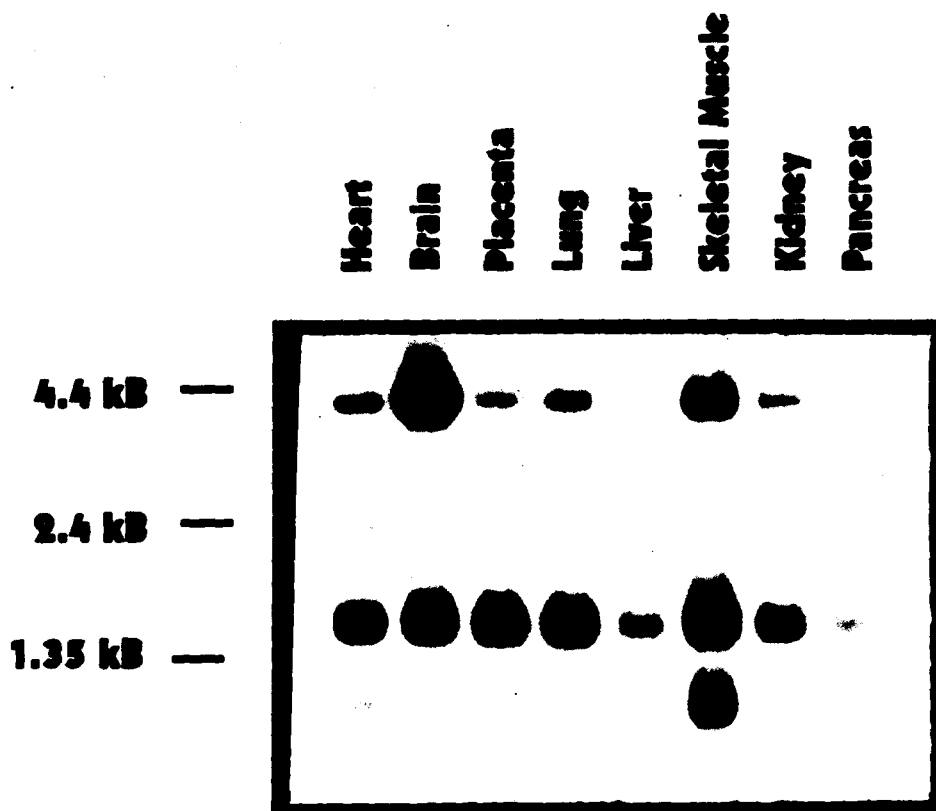Figure 20:   CALM 1 on Human Multiple Tissue Northern Blot.   The CALM 1 gene is expressed in all tissues examined with the 4.0 kb transcript expressed at the highest level in brain.   The 4.0 kb and 1.4 kb transcripts are known to originate from the CALM 1 gene (Rhyner *et al.*, 1994) but the 0.8 kb transcript is suspected to result from cross-hybridization of the CALM 1 probe with either the CALM2 or CALM 3 transcript.

identical situation existed in the human *CALM 1* gene and Rhyner *et al.* (1994) have recently shown that the 4.0 kb and 1.5 kb transcripts are generated by the use of alternative polyadenylation sites in the human *CALM 1* gene. The presence of the 0.8 kb transcript is thought to be due to cross hybridization CALM 1 probe with either the *CALM 2* or *CALM 3* mRNA. It is observed that the 4.0 kb transcript is produced in greatest amounts in brain. This could suggest a function for the differing 3' ends in tissue specific regulation possibly by modulating transcript stability.

B.   Random RACE Cloning of (CAG)$_N$ Containing cDNA Fragments

Recognizing the limitation of 3' RACE to only clone fragments of cDNA's that have a (CAG)$_N$ repeat located near the 3' end an attempt was made to develop a methodology that would allow the cloning of cDNA fragments that contained (CAG)$_N$ repeats located in any region of a transcript. The method that was eventually developed utilized a similar rapid amplification of cDNA ends (RACE) approach with the adaptation of utilizing a random oligo for reverse transcription. Jurkat cell line total RNA was utilized for this set of experiments as it allowed for a good source of high quality RNA and would not seem to hinder the isolation of potential disease genes as all of the known diseases of trinucleotide repeat expansion involve genes that are expressed in

TABLE 5: JURKAT RANDOM RACE CLONES

Clone	Repeats	GenBank Search	Clone	Repeats	GenBank Search
JRR3	6	c-myc	JRR44	NA	No match
JRR4	NA	No match	JRR46	NA	No match
JRR5	NA	No match	JRR48	NA	No match
JRR10	25-42	TBP	JRR49	NA	No match
JRR12	NA	No match	JRR50	NA	No match
JRR13	NA	MEF2	JRR52	NA	No match
JRR14	~20	No match	JRR53	NA	No match
JRR15	3	est27h12	JRR54	NA	No match
JRR16	NA	No match	JRR55	NA	No match
JRR17	9	Nopp 140	JRR60	NA	No match
JRR21	NA	No match	JRR62	NA	No match
JRR22	8	CALM 1	JRR64	5	sts ut1532
JRR24	NA	No match	JRR65	NA	No match
JRR29	NA	No match	JRR66	NA	No match
JRR30	NA	BarH1	JRR68	NA	RNA Helicase
JRR31	NA	No match	JRR72	NA	No match
JRR32	NA	No match	JRR78	NA	No match
JRR33	NA	No match	JRR80	NA	No match
JRR35	NA	No match	JRR88	NA	No match
JRR36	NA	No match	JRR92	NA	No match
JRR38	NA	No match	JRR95	NA	No match
JRR40	2	28S rRNA	JRR96	NA	No match

Table 5: Random RACE Clones. The clones isolated by carrying out the Random RACE technique are listed. All clones were subjected to DNA sequence analysis and the resulting sequence was utilized in GenBank search. The number of repeats reported was established for clones that had matches in GenBank. Those listed as NA contained the 8 repeats from the CAG8 oligo.

a variety of tissues. Over 100 clones, referred to as JRR (Jurkat Random RACE), were initially isolated and subject to sequence tag analysis. The clones isolated in this experiment are summarized in Table 5. Out of an initial pool of 100 clones the 44 listed in Table 5 were found to be informative. The remaining clones are comprised of duplicates and a small number of cloning artifacts. Additionally, a group of approximately 30 clones is still in the process of being analyzed. Notably, several clones were isolated that were shown to be identical to previously isolated cDNA's that are known to contain $(CAG)_N$ repeats. This observation served as confirmation that the technique was operating as designed. Additionally, the *CALM 1* cDNA was isolated by this approach, confirming the 3' RACE results. More detailed description of several of the clones will be covered in the following sections.

1. __Clone JRR3__

Clone JRR3 was isolated and subject to sequence analysis. GenBank search revealed that the 82 bp clone was identical to a region of exon 2 of the *c-myc* proto-oncogene. This region of *c-myc* has been shown to contain a stretch of six CAG repeats that code for a stretch of glutamines (Watson et al., 1983) It is interesting that our eight repeat primer was able to amplify from a six repeat region and may indicate polymorphism at this site.

2.   Clone JRR10

Clone JRR10 was subjected to DNA sequence analysis with the resulting sequence utilized in a GenBank search. The results of this search revealed that the 230 bp insert of JRR10 was identical to a region in the 5' portion of the TATA binding protein (TBP) cDNA (Hoffmann et al., 1990). The TBP protein is known to contain a large polyglutamine region that is encoded by this CAG repeat region and as indicated in Table 5 this region of the *TBP* gene is known to be polymorphic showing a range of 25 to 42 repeats (Polymeropoulos *et al.*, 1991; Gostout *et al.*, 1993). Further analysis of the JRR clones revealed that a separate clone isolated (JRR47) is identical to JRR10.

3.   Clone JRR15

Clone JRR15 was subjected to sequence analysis with the resulting sequence utilized in a GenBank search. The results of the search revealed that JRR15 is identical to expressed sequence tag (EST) 27H12 (GenBank accession Z15569). Interestingly, the sequence of EST 27H12 showed only three CAG repeats. This result could be accounted for if this repeat region is polymorphic and the template cDNA from Jurkat cells contained a larger number of repeats.

4.   Clone JRR17

Clone JRR17 was subjected to DNA sequence analysis and GenBank search with this sequence revealed that the clone was identical to a human sequenced open reading frame (GenBank accession 94288).   This open reading frame apparently is the human homologue to the rat *Nopp140* gene and contains 9 repeats of CAG interrupted with a TAG.   This region is in the 5' portion of the cDNA and is predicted to code for a stretch of poly-serine.   In the rat, the Nopp140 protein has been shown to be a nucleolar phosphoprotein involved in nuclear protein import (Meier and Blobel, 1990).

5.   **Clone JRR30**

Clone JRR30 was analyzed by DNA sequencing and the resulting sequence was utilized in a GenBank search.   The results of the search showed that a region of the clone was similar to the 5' translated region of the *D. melanogaster BarH1* homeobox gene. Subsequent translation of one of the reading frames of JRR30 results in a peptide which has three regions of similarity to the

```
BarH1  1    M K D S M S I L T Q T P S E P N A A H P Q L H H H L S T  L Q
JRR30  1    - - - - - - - - - - - - - - - A A A L P Q M P L F E N F  Y F

BarH1  31   - - Q Q H H Q H H L H Y G L Q P A A V A H S I H S T T T  M S
JRR30  16   M P Q Q P S Q Q P Q D F G L Q P A G P L G Q S H L A H H  S M

BarH1  59   S G G
JRR30  46   A P Y
```

Figure 21: Comparison of a region of JRR30 to the *D. melanogaster*
BarH1 amino acid sequence. A BLAST search (Altschul *et al.*, 1990)
discovered a region of similarity between clone JRR30 and the
Drosophila homeobox gene *BarH1*. A comparison of a putative reading
frame of JRR30 with the N-terminus of the BarH1 protein is shown.
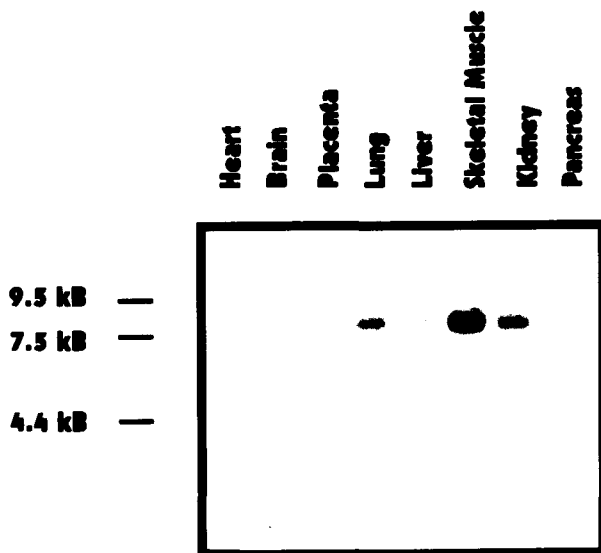The boxes indicate areas with a high degree of similarity.

Figure 22: JRR30 on Human Multiple Tissue Northern Blot. A Human Multiple Tissue Northern Blot was probed with the JRR30 cDNA. The results of this experiment are shown and indicate that the clone detects an ~8.0 kb transcript that is produced in all tissues examined.

N-terminus of the *BarH1* gene product (Figure 21). Over a 31 amino acid region the two sequences are 38% identical and 52% similar. This region of similarity is small but may indicate the conservation of a particular domain between humans and *D. melanogaster.* However, without the entire JRR30 cDNA sequence it is not possible to conclusively predict the reading frame for this fragment.

Expression analysis of the JRR30 gene was carried out by hybridizing the radioactively labeled fragment to a Human Multiple Tissue Northern blot (Clontech). The JRR30 fragment detects an ~8.0 kb transcript that is expressed in all tissues examined. Additionally, there were several smaller weakly hybridizing transcripts that may be indicative of a gene family.

6.   **Clone JRR64**

Clone JRR64 was isolated and subjected to DNA sequence analysis. GenBank search with this sequence revealed that clone JRR64 is identical to the short transcribed sequence (STS) UT1532. STS UT1532 has been shown to map to chromosome 15 and it contains an imperfect repeat of six CAG's interrupted with the dinucleotide sequence CG (GenBank accession L16404). Aside from the chromosomal assignment nothing is known regarding STS UT1532.

7. **Novel JRR Clones**

The results of GenBank searches with all of the JRR clones analyzed are shown in Table 5. The GenBank search results indicate that of the 44 clones listed, 34 show no significant similarity to any sequences in GenBank. This collection of 34 novel $(CAG)_N$ containing cDNA fragments along with several of the novel clones described above provide a pool of useful molecular reagents. These clones provide the basis for future experiments aimed at the molecular dissection of trinucleotide repeat expansion disorders.

CHAPTER V

DISCUSSION

It is now apparent that trinucleotide repeat expansion is a major type of mutagenesis leading to human diseases. The expansion of these GC rich repeats located within transcribed sequences is now known to lead to two separate classes of diseases. Class I disorders are the late onset neurological disorders caused by expansion of $(CAG)_N$ repeats that are in translated regions of the respective genes. Class II disorders are caused by expansion of GC rich repeats present in untranslated regions of the 3 respective genes.

The experimental results described in this dissertation clearly show that rapid amplification of cDNA ends (RACE) is a useful technique for the isolation of trinucleotide repeat containing cDNA's. The 3' RACE experiments indicate that there appear to be a small subset of mRNA's that contain a $(CAG)_N$ repeat located near the 3' end. The CAG4 oligo experiment indicates that there are a number of novel transcripts that would appear to contain $(CAG)_N$ repeats within 3' untranslated regions. As an example, the clone CAG4-6 is identical to the 3' 381 bp of the

human $G_{\alpha s}$ cDNA. Also, the length of the sequence of clone CAG4-3 is sufficient to contain the repeat within a translated region of the respective transcript. The size of the remaining clones would predict that the repeat resides in the 3' UTR of the respective transcripts.

The CAG8 3' RACE experiments provided nine different clones. The sequences of six clones did not identify any previously isolated genes, while CAG8-5, CAG8-16, and hbCAG8-54 were found to encode SRP14, p9-ssbp, and calmodulin 1, respectively. As a result of this work the coding sequences for CAG8-6 and CAG8-16 (p9-ssbp) have been elucidated and provide the basis for further study into their respective cellular functions.

Clone CAG8-6 shows no strong similarity to any known genes yet the protein contains over 10% arginine and serine, respectively. Similarly, the S/R proteins that function in RNA splicing are known to contain high amounts of arginine and serine. However, the regions are generally made up of Arg-Ser dipeptide repeats, which is not observed in CAG8-6. Additionally, CAG8-6 does not appear to show a consensus RNA recognition motif which is another characteristic of S/R proteins. From the available data it would appear that CAG8-6 does not belong to the S/R family of proteins.

The other interesting observation of CAG8-6 is the stability of the CAG repeat. On an evolutionary basis one may speculate

that the fact that the repeat does not code for the hydrophilic polyglutamine but instead the more hydrophobic serine leads to selection against variability within the repeat. This hypothesis is in agreement with the suggestion of Green and Wang (1994) that $(CAG)_N$ sequences are an evolutionary mechanism for proteins to add amino acids.

Green and Wang (1994) have proposed that $(CAG)_N$ trinucleotide repeats function as an evolutionary mechanism for protein sequences to add amino acids. It is postulated that stretches of polyglutamine could be added to proteins and allowing for normal spontaneous mutations the region could mutate under selective pressure to other sequences. Green and Wang (1994) state that protein structure is better able to withstand substitution by hydrophilic amino acids than hydrophobic residues and would thereby generate a selective advantage for $(CAG)_N$ sequences to code for polyglutamine. Consequently, a corollary of this hypothesis would predict the existence of selective pressure against the polymorphic variation of hydrophobic amino acids. Additionally, once a stretch of polyglutamine has been inserted into a protein it could expand by the same mechanism that is functioning in the diseases. Therefore, proteins would not only be able to try out new sequences they would also be able to increase in size. The critical point in this hypothesis is that the expansion always occurs in multiples of three and would therefore retain the open reading frame.

Clone CAG8-16 encodes the human homologue of a rodent single-stranded DNA binding protein, as revealed by GenBank search. Although CAG8-16 codes for a protein of known biochemical function (Ballard et al., 1988) the cellular role of the ssbp is unclear. A plausible hypothesis is that the protein is involved in DNA synthesis which would then include cellular division, DNA repair, and recombinational repair. However, all of this is speculation without the aid of further experimental data. Another question to be answered is whether the human p9-ssbp protein is proteolytically processed in a manner similar to the mouse protein. This proteolytic processing provides a point of cellular regulation whereby the protein is synthesized in an inactive form and could potentially be sequestered until it is needed. Finally, similar to the CAG8-6 protein, the $(CAG)_6$ repeat present in the p9-ssbp is coded as serine and would therefore be predicted to show little or no polymorphism among the population due to the inability of the protein structure to withstand variable tracts of serine.

Another result of the CAG8 3' RACE experiments as mentioned above was the isolation of the *CALM 1* cDNA. Initially, this was a surprising result due to the absence of a $(CAG)_N$ repeat in the reported *CALM 1* sequence. Now that this finding has been confirmed by the cloning of *CALM 1* gene (Rhyner et al., 1994) and by the re-isolation of a fragment of *CALM 1* by the Random RACE method the next logical question is what is the nature of the CAG

repeat in the normal population. Strehler and coworkers are currently addressing this question (personal communication). This case may provide an interesting comparison since the CAG repeat in CALM 1 is located within the 5' UTR while the CAG repeats of class I disorders are all located within translated regions. Furthermore, a larger sampling of genes containing CAG repeats in untranslated regions should be examined to investigate any possible differences in stability of the sequence depending on its location within a gene.

The Random RACE technique was the last set of experiments described and the results illustrate that the technique is well suited to the isolation of trinucleotide repeat containing cDNA fragments. Another possible benefit of the technique is the possible application to experimental situations where only a partial sequence of a unique clone is known and obtaining a larger cDNA clone is the goal. Although no direct evidence of the technique being applicable to such situations is presented here it would seem to be a technically feasible application of the methodology.

The majority of the cDNA clones isolated by the Random RACE technique are novel and therefore generate a library of clones to be utilized in the examination of genetic diseases potentially caused by the expansion of $(CAG)_N$ sequences in translated regions. Several of the clones, in particular JRR30, show some slight degree of similarity to known genes. The small region of

similarity between JRR30 and the *D. melanogaster BarH1* gene may indicate the conservation of a functionally important peptide domain. *BarH1* is a known homoebox gene that is required for proper eye development in the fruit fly. The region of similarity to JRR30 does not include the homeobox domain and the presence of such a domain in the JRR30 cDNA must await the isolation of this cDNA. Furthermore, the cloning of this cDNA would provide the number of CAG repeats present in the gene and give the necessary sequences to synthesize oligonucleotides to examine the possible polymorphic nature of the repeat.

The progress on diseases of trinucleotide expansion has been a surprising result of modern molecular biology. SBMA was described in 1991 as the first disorder and the other six have been discovered in the last three years. The association of these types of disorders with clinical variability and genetic anticipation now provides a genetic basis for screening new disorders to examine them for the possibility of trinucleotide repeat expansion. This type of approach has already yielded the molecular dissection of both SCA1 and DRPLA (Orr *et al.*, 1993; Koide *et al.*, 1994; Nagafuchi *et al.*, 1994). The combined effort of Li *et al.* (1993), Koide and coworkers (1994), and Nagafuchi and coworkers (1994) illustrates the potential benefits of random cloning of cDNA's containing trinucleotide repeats. The clones isolated as a result of this dissertation research provide a set of cDNA fragments that contain $(CAG)_N$ repeats that have a high

probability of being present in translated regions. Given the involvement of the expansion of $(CAG)_N$ sequences in late onset neurological disorders it may be beneficial to screen similar late onset neurological diseases with all of the novel CAG8 and JRR clones.

Possibilities for this type of application already exist as Zoghbi (personal communication) has reported the existence of numerous late onset ataxias which do not fit into the categories of SCA1, Machado-Joseph disease (SCA2), or SCA3. These molecularly undefined disorders show clinical variability and anticipation yet are not present in large enough pedigrees to allow for linkage analysis. Furthermore, anticipation has been reported in familial schizophrenia and may indicate that expansion of trinucleotide repeats is involved as a causative factor (Bassett and Honer, 1994). Overall, it seems a very plausible hypothesis that more disorders caused by expansion of trinucleotide repeats are going to be described in the future and it is expected that some of the cDNA fragments isolated during the course of this work will provide useful molecular reagents for analysis of these disorders.

Mechanistically, one of the more interesting and challenging tenets of trinucleotide repeat expansion is that it has only been observed in humans. The polymorphic variation of $(CAG)_N$ repeats has been observed in diverse species as far down the evolutionary ladder as yeast (Tautz, 1989). However, transgenic mice carrying

an expanded allele of the androgen receptor 1) show no phenotype similar to SBMA and 2) the expanded allele has been observed to be stable through 40 meiosis (K. Fishbeck, personal communication). This point indicates that humans have some *trans* acting factor which is involved in the expansion of these repeats.

In a related research direction work in both DM and FRAXA has indicated that by haplotype analysis both disorders show a strong founder effect (Imbert *et al.*, 1993; Neville *et al.*, 1994; Smits *et al.*, 1993). Similarly, analysis of new mutations in HD has revealed an apparent premutation range which may be made up of a subset of chromosomes that are predisposed to expansion (Goldberg *et al.*, 1993; Myers *et al.*, 1993). This work indicates that in addition to required *trans* acting factors that specific diseases may have predisposing cis acting regions that lead to instability. This leads to the speculation that expansion of a repetitive element into the disease range may require some yet unknown interaction between these *cis* and *trans* factors.

In a related matter it appears for the two classes of diseases there may be a sex of parent influence on the instability of the repetitive element. Disease alleles of all of the class I disorders have been demonstrated to have a higher degree of instability when paternally inherited (Biancalana *et al.*, 1992; Chung *et al.*, 1993; Telenius *et al.*, 1993; Koide *et al.*, 1994). This is in contrast to the class II disorders where investigations of DM and FRAXA have shown enhanced meiotic instability as a

result of maternal inheritance of a disease allele (Tsilfidis *et al.*, 1992; Lavedan *et al.*, 1993; Heitz *et al.*, 1992). This data may suggest a different mechanism of expansion depending on the sequence and location of the trinucleotide repeat.

The search for the enzymatic mechanism of the expansion of trinucleotide repeats is intensely studied. Several hypotheses have been put forth to account for expansion of trinucleotide repeats including unequal crossing over (unequal reciprocal exchange of information) (Sturtevant, 1925), gene conversion (non-reciprocal exchange of information), and replicative slippage (O'Hoy *et al.*, 1993; *Fu et al.*, 1991; Strand *et al.*, 1993; Kunkel, 1993; and Figure 23). The hypothesis that the expansion is the result of unequal crossing over was originally proposed as an explanation of the polymorphic nature of dinucleotide repeats (Levinson and Gutman, 1987). However, analysis of the genomic regions surrounding expanded trinucleotide repeats has revealed no alteration of flanking markers making a recombinational mechanism such as unequal crossing over unlikely (Fu *et al.*, 1992; HD Collaborative Research Group, 1993). The involvement of a possible gene conversion mechanism has been given some attention following the work of O'Hoy *et al.* (1993) who reported a case where a daughter of a DM positive male showed two markers within a 7.2 kb region that were derived from the father's unaffected chromosome while two different markers within this 7.2 kb region were from the affected chromosome. Additional marker analysis

showed that the paternally derived chromosome was the abnormal DM chromosome. Interestingly, the child's DM gene contained only 13 CTG repeats that had apparently been converted from the father's normal chromosome. Given that this report was a rare case where a contraction of repeat number is observed and due to the alteration of flanking markers it would seem that this putative gene conversion event is best described as an isolated case rather than a general mechanism for the expansion of trinucleotide repeats.

The mechanism that is most often referred to as the probable cause of trinucleotide repeat expansion is that of replicative slippage. This is best described as a slippage of either the newly synthesized or the template strand during polymerization that results in expansion or contraction of the replicated sequence depending on the strand that has slipped (Figure 23). This model provides for a sound explanation for small expansions and for the polymorphic nature of the repeats in the normal population however, further description is necessary to explain the observance of large expansions in both FRAXA and DM.

One adaptation of the theory has been put forth by Richards and Sutherland (1994) that is based on the repeat region being coincident with an Okazaki fragment during replication. The authors propose that as repeat lengths pass a threshold they would move into a size range larger than the average size of an Okazaki fragment and would thereby allow for the fragment to be unanchored to any unique sequence. This would result in the unregulated

slippage during replication that could explain the observation of very large expansions.

Similarly, Fu *et al.* (1991) have proposed that the GC rich regions of these repeats leads to replicative termination and reinitiation within the repeat sequences. This would then give multiple incomplete strands the opportunity to mispair before the replicative process is complete. In support of the replicative slippage model Schlötterer and Tautz (1992) have shown *in vitro* reactions utilizing a 15 bp repeat template and 9 bp repeat primer with *E. coli* DNA polymerase I gives rise to larger products indicative of replicative slippage. Perhaps even more interesting is the finding that only those template primer pairs that were at least 2/3 GC were able to realign in an integral of three. Utilizing template:primer pairs that were 2/3 AT resulted in slippage by integral values of two, three and four. This data would suggest that expansion of trinucleotide repeats within translated regions will be confined to GC rich repeats due to the fact that slippage of AT rich repeats could result in frameshift mutations. Expansion of AT rich sequences located in untranslated sequences could still be a possibility.

Finally, Green (1993) has proposed an attractive mechanism to specifically explain the cellular pathology of the four class I disorders whose $(CAG)_N$ repeats code for polyglutamine. Green
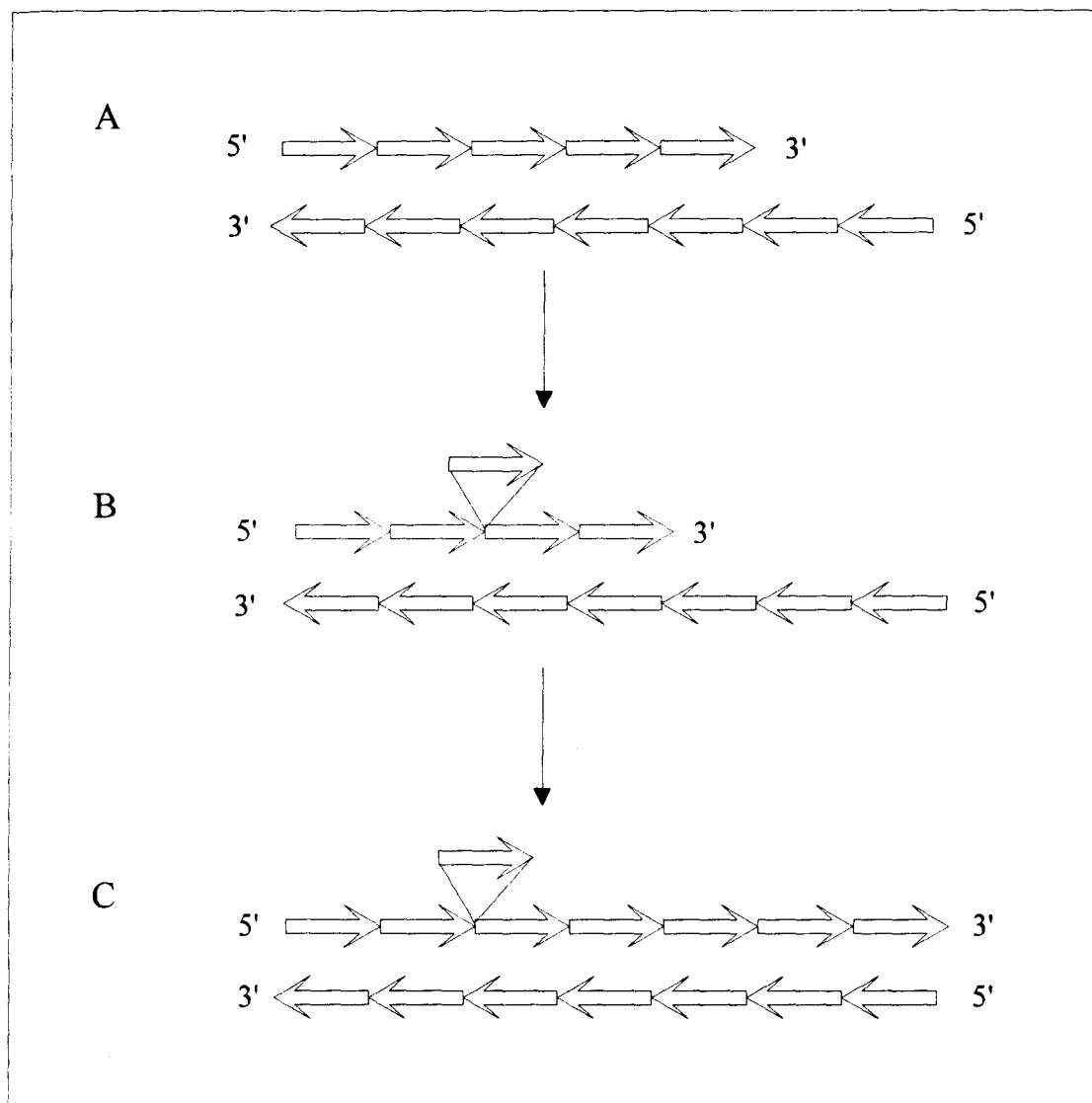
Figure 23: Schematic of Replicative Slippage. The diagram illustrates the model proposed for replicative slippage. The arrows represent a trinucleotide repeat unit and step A is the starting template. In step B the primer strand has slipped by one repeat unit followed by subsequent replication in step C. If the primer strand is not repaired following this slippage/replication cycle in the next replicative cycle one of the daughter products will have increase by one repeat unit. It is assumed that the slippage could occur in multiple repeat units.

(1993) points out that the involucrin protein contains a high percentage of CAG repeats or similar codons that have a different base in the third position. Consequently, the protein contains one region of 18 glutamines and another of six glutamines. During keratinocyte differentiation the involucrin protein is cross-linked by a glutamine-lysine linkage to other cellular proteins by a transglutaminase. Green proposes that the protein products of the four diseases HD, SBMA, SCA1, and DRPLA could be poor substrates for transglutaminase, even in normal individuals. As the CAG repeats expand in the four class I disorders, the proteins could become improved substrates for the transglutaminase, which is known to be expressed in neural cells. This would allow the disease proteins to become cross-linked to lysine donor proteins and generate aggregates with proteins that do not generally interact with the respective disease proteins. Although the aggregates are believed to be proteolyzed, the dipeptide cross-link can not be degraded and would accumulate within the cell. Additionally, several possibilities are discussed for the neuronal specificity. These include 1) transglutaminases are known to be important in synaptic transmission, 2) neurons are less able to degrade transglutaminase aggregates, 3) there is no renewal of neurons to compensate for cell death, and 4) the transglutaminase is activated by Ca^{2+} which is present in high levels in neurons. The hypothesis presented by Green (1993) accounts for many of the

characteristics of the pathology of the four class I disorders including the gradual progression/late onset and neuronal specificity. Furthermore, Green indicates that this hypothesis could be tested utilizing immunocytochemistry with antibodies that recognize the four class I proteins and the dipeptide linkage generated by the transglutaminase.

Overall, it is now known that trinucleotide repeat expansion is a unique and important class of mutagenesis leading to human disease. Given the short time in which the seven known disorders have been described and the existence of other molecularly undefined disorders with similar genetic characteristics, it is expected that many more diseases will be associated with trinucleotide repeat expansion. Given this, it is hoped that the cDNA fragments isolated as a result of this dissertation research will contribute to the molecular characterization of a number of these disorders.

LITERATURE CITED

Aaltonen, L.A., P. Peltomäki, F.S. Leach, P. Sistonen, L. Pylkkänen, J-P. Mecklin, H. Järvinen, S.M. Powell, J. Jin, S. R. Hamilton, G.M. Petersen, K.W. Kinzler, B. Vogelstein, and A. de la Chapelle. 1993. Clues to the pathogenesis of familial colorectal cancer. Science **260**: 812-816.

Abitbol, M., C. Menini, A-L. Delezoide, T. Rhyner, M. Vekemans, and J. Mallet. 1993. Nucleus basalis magnocellularis and hippocampus are the major sites of FMR-1 expression in the human fetal brain. Nat. Genet. **4**: 147-153.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403-410.

Ambrose, C.M., M.P. Duyao, G. Barnes, G.P. Bates, C.S. Lin, J. Srinidhi, S. Baxendale, H. Hummerich, H. Lehrach, M. Altherr, J. Wasmuth, A. Buckler, D. Church, D. Housman, M. Berks, G. Micklem, R. Durbin, A. Dodge, A. Read, J. Gusella, and M.E. MacDonald. 1993. Structure and expression of the Huntington's disease gene: Evidence against simple inactivation due to an expanded CAG repeat. Som. Cell and Mol. Genet. **20**: 27-38.

Ashizawa, T., M Anvret, M. Baiget, J.M. Barceló, H. Brunner, A.M. Cobo, B. Dallapiccola, R.G. Fenwick Jr., U. Grandell, H. Harley, C. Junien, M.C. Koch, R.G. Korneluk, C. Lavedan, T. Miki, J.C. Mulley, A. López de Munain, G. Novelli, A.D. Roses, W.K. Seltzer, D.J. Shaw, H. Smeets, G.R. Sutherland, H. Yamagata, and P.S. Harper. 1994. Characteristics of intergenerational contractions fo the CTG repeat in myotonic dystrophy. Am. J. Hum. Genet. **54**: 414-423.

Ashley, C., Jr., K.D. Wilkinson, D. Reines, and S.T. Warren. 1993. FMR1 protein: Conserved RNP family domains and selective RNA binding. Science **262**: 563-566.

Ashley, C.T., J.S. Sutcliffe, C.B. Kunst, H.A. Leiner, E.E. Eichler, D.L. Nelson, and S.T. Warren. 1993. Human and murine FMR-1: Alternative splicing and translational initiation downstream of the CGG-repeat. Nat. Genet. **4**: 244-251.

Ballard, D.W., W.M. Philbrick, and A.L.M. Bothwell. 1988. Identification of a novel 9-kDa polypeptide from nuclear extracts. J. Biol. Chem. **263**: 8450-8457.

Banfi, S., A. Servadio, M-Y. Chung, T.J. Kwiatkowski Jr., A.E. McCall, L.A. Duvick, Y. Shen, E.J. Roth, H.T. Orr, and H.Y. Zoghbi. 1994. Identification and characterization of the gene causing type 1 spinocerebellar ataxia. Nat. Genet. **7**: 513-520.

Banfi, S., M-Y. Chung, T.J. Kwiatkowski Jr., L.P. Ranum, A.E. McCall, A.C. Chinault, H.T. Orr, and H.Y. Zoghbi. 1993. Mapping and cloning of the critical region for the spinocerebellar ataxia type 1 gene (SCA1) in a yeast artificial chromosome contig spanning 1.2 Mb. Genomics **18**: 627-635.

Barron, L.H., A. Rae, S. Holloway, D.J.H. Brock, and J.P. Warner. 1993. A single allele from the polymorphic CCG rich sequence immediately 3' to the unstable CAG trinucleotide in the IT15 cDNA shows almost complete disequilibrium with Huntington's disease chromosomes in the Scottish population. Hum. Mol. Genet. **3**: 173-175.

Bassett, A., and W.G. Honer. 1994. Evidence for anticipation in schizophrenia. Am. J. Hum. Genet. **54**: 864-870.

Biancalana, V., F. Serville, J. Pommier, J. Julien, A. Hanauer, and J.L. Mandel. 1992. Moderate instability of the trinucleotide repeat in spino bulbar muscular atrophy. Hum. Mol. Genet. **1**: 255-258.

Bronner, C.E., S.M. Baker, P.T. Morrison, G. Warren, L.G. Smith, M.K. Lescoe, M. Kane, C. Earabino, J. Lipford, and A. Lindblom. 1994. Mutation in the DNA mismatch repair gene homologue hMLH1 is assocaiated with hereditary non-polypsis colon cancer. 1994. Nature **368**: 258-261.

Brook, J.D., M.E. McCurrah, H.G. Harley, A.J. Buckler, D. Church, H. Aburatani, K. Hunter, V.P. Stanton, J-P. Thirion, T. Hudson, R. Sohn, B. Zemelman, R.G. Snell, S.A. Rundle, S. Crow, J. Davies, P. Shelbourne, J. Buxton, C. Jones, V. Juvonen, K. Johnson, P.S. Harper, D.J. Shaw, and D.E. Housman. 1992. Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat in the 3' end of a transcript encoding a protein kinase family member. Cell **68**: 799-808.

Brunner, H.G., G. Jansen, W. Nillesen, M.R. Nelen, C.E.M. de Die, C.J. Höweler, B.A. van Oost, B. Wieringa, H-H. Ropers, and H.J.M.

Smeets. 1993. Brief report: Reverse mutation in myotonic dystrophy. N. Engl. J. Med. **328**: 476-480.

Bryer, A., R.W. Martell, E.D. du Toit, and P. Beighton. 1992. Adult onset spinocerebellar ataxia linked to HLA in a south african kindred of mixed ancestry. Tissue Antigens **40**: 11-115.
Burke, J.R., M.S. Wingfield, K.E. Lewis, A.D. Roses, J.E. Lee, C. Hulette, M.A. Pericak-Vance, and J.M. Vance. 1994. The haw river syndrome: Dentatorubropallidoluysian atrophy (DRPLA) in an african-american family. Nat. Genet. **7**: 521-524.

Carango, P., J.E. Noble, H.G. Marks, and V. L. Funanage. 1993. Absence of myotonic dystrophy protein kinase (DMPK) mRNA as a result of a triplet repeat expansion in myotonic dystrophy. Genomics **18**: 340-348.

Carlock, L., K. Gutridge, and T. Vo. 1994. A *Sau*3A polymorphism in the 5' end of the IT15 gene that nonrandomly segregates with the Huntington disease trinucleotide expansion. Hum. Genet. **93**: 457-459.

Cheng, K.C. and L.A. Loeb. 1993. Genomic instability and tumor progression: Mechanistic considerations. Adv. Cancer Res. **60**: 121-156.

Chomczynski, P. and N. Sacchi. 1989. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**: 156-159.

Chung, M-Y., L.P.W. Ranum, L.A. Duvick, A. Servadio, H.Y. Zoghbi, and H.T Orr. 1993. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type 1. Nat. Genet. 5: 254-258.

Cohen, D., I. Chumakov, and J. Weissenbach. 1993. A first-generation physical map of the human genome. Nature **366**: 698-700.

Currier, R.D., G. Glover, J.F. Jackson, and A.C. Tipton. 1972. Spinocerebellar ataxia: Study of a large kindred. I. General information and genetics. Neurol. **22**: 1040-1043.

DeBoulle, K., A.J.M.H. Verkerk, E. Reyniers, L. Vits, J. Hendrickx, B. Van Roy, F. Van Den Bos, E. deGraaf, B.A. Oostra, and P.J. Williams. 1993. A point mutation in the FMR-1 gene associated with mental retardation. Nat. Genet. **3**: 31-35.

Devys, D., Biancalana, V., F. Rousseau, J. Boué, J.L. Mandel, and I. Oberlé. 1992. Analysis of full frafile X mutations in fetal tissues and monozygotic twins indicates that abnormal methylation

and somatic heterogeneity are established early in development. Am. J. Med. Genet. **43**: 208-216.

Dietrich, A., P. Kioschis, A.P. Monaco, B. Gross, B. Korn, S.V. Wiliiams, D. Sheer, D. Heitz, I. Oberle, D. Toniolo, S.T. Warren, H. Lehrach, and A. Poutska. 1991. Molecular cloning and analysis of the fragile X region in man. Nucl. Acids. Res. **19**: 2567-2572.

Doolitle, R. 1986. Of URF's and ORF's: A Primer on How to Analyze Derived Amino Acid Sequences. University Science Books, Mill Valley, CA.

Edwards, A., A. Civitello, H.A. Hammond, and C.T. Caskey. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. **49**: 746-756.

Edwards, A., H.A. Hammonds, L. Jin, C.T. Caskey, and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics **12**: 241-253.

Etongué-Mayer, P., R. Faure, J-P. Bouchard, M-C. Thibault, and J. Puymirat. 1994. The myotonin-protein kinase phosphorylates tyrosine residues in normal human skeletal muscle. Biochem. Biophys. Res. Comm. **199**: 89-92.

Feinberg, A.P. and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction fragments to high specific activity. Anal. Biochem. **132**: 6-13.

Fishbeck, K.H., V. Ionasescu, A.W. Ritter, R. Ionasescu, K. Davies, S. Ball, P. Bosch, T. Burns, I. Hausmanowa-Petrusewicz, J. Borkowska, S.P. Ringel, and L.S. Stern. 1986. Localization of the gene for X-linked spinal muscular atrophy. Neurol. **36**: 1595-1598.

Fishel, R., M.K. Lescoe, M.R. Rao, N.G. Copeland, N.A. Jenkins, J. Garber, M. Kane, and R. Kolodner. 1993. The human mutator gene homolog and its association with hereditary nonpolypsis colon cancer. Cell **75**: 1027-38.

Fleischer, B. 1918. Über myotonische dystrophie mit katarakt: Eine hereditäre, familiäre degeneration. Archiv. für Ophthalmologie **96**: 91-133.

Flynn, G.A., M.C. Hirst, S.J.L. Knight, J.N. Macpherson, J.C.K. Barber, A.V. Flannery, K.E. Davies, V.J. Buckle. 1993. Identification of the FRAXE fragile site in two families ascertained for X linked mental retardation. J. Med. Genet. 30: 97-100.

Fu, Y-H., A. Pizzuti, R.G. Fenwick Jr., J. King, S. Rajnarayan, P.W. Dunne, J. Dubel, G.A. Nasser, T. Ashizawa, P. DeJong, B. Wieringa, R. Korneluk, M.B. Perryman, H.F. Epstein, and C.T. Caskey. 1992. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science **255**: 1256-1258.

Fu, Y-H., D.L. Friedman, S. Richards, J.A. Pearlman, R.A. Gibbs, A. Pizzuti, T. Ashizawa, M.B. Perryman, G. Scarlato, R.G. Fenwick Jr., and C.T. Caskey. 1993. Decreased expression of myotonin-protein kinase messenger RNA and protein in adult form of myotonic dystrophy. Science **260**: 235-238.

Fu, Y-H., D.P.A. Kuhl, A. Pizzuti, M. Pieretti, J.S. Sutcliffe, S. Pichards, A.J.M.H. Verkerk, J.J.A. Holden, R.G. Fenwick Jr., S.T. Warren, B.A. Oostra, D.L. Nelson, and C.T. Caskey. 1991. Variation of the CGG repeat at the Fragile X site results in genetic instability: resolution of the Sherman paradox. Cell **67**: 1047-1058.

Gerber, H.P., K. Seipel, O. Georgiev, M. Höfferer, M. Hug, S. Rusconi, and W. Schaffner. 1994. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science **263**: 808-811.

Gieser, L., A. Swaroop. 1992. Expressed sequence tags and chromosomal localization of cDNA clones from a subtracted retinal pigment epithelium library. Genomics 13: 873-876.

Goldberg, Y.P., B. Kremer, S.E. Andrew, J. Theilmann, R.K. Graham, F. Squitieri, H. Telenius, S. Adam, A. Sajoo, E. Starr, A. Heiberg, G. Wolff, and M. Hayden. 1993. Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex origin effects. Nature Genet. **5**: 174-179.

Gostout, B., Q. Liu, and S.S. Sommer. 1993. "Cryptic" repeating triplets of purines and pyrimidines (cRRY(i)) are frequent and polymorphic: Analysis of coding (cRRY(i)) in proopiomelanicortin (POMC) and TATA-binding protein (TBP) genes. Am. J. Hum. Genet. **52**: 1182-1190.

Grabowski, D.T., J.P. Carney, and M.R. Kelley. 1991. A *Drosophila* gene containing the repetitive element *opa* is exclusively expressed in adult male abdomen. Nucl. Acids. Res. **19**: 1709.

Green, H. 1993. Human genetic disease due to codon reiteration: Relationship to an evolutionary mechansim. Cell **74**: 955-956.

Green, H. and N. Wang. 1994. Codon reiteration and the evolution of proteins. Proc. Natl. Acad. Sci. 91 (1994) 4298-4302.

Greenfeld, J.G. 1954. The spino-cerebellar degenerations. Charles C. Thomas, Springfield, Il.

Gusella, J.F., M.E. MacDonald, C.M. Ambrose, and M.P. Duyao. 1993. Molecular genetics of huntington's disease. Arch. Neurol. **50**: 1157-1163.

Gusella, J.F., M.E. MacDonald, C.M. Ambrose, M.P. Duyao. 1993. Molecular genetics of Huntington's disease. Arch. Neurol. **50**: 1157-1163.

Gusella, J.F., N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Tanzi, P.C. Watkins, K. Ottina, M. R. Wallace, A.Y. Sakaguchi, A.B. Young, I. Shoulson, E. Bonilla, and J.B. Martin. A polymorphic DNA marker genetically linked to Huntington's disease. 1983. Nature **306**: 234-238.

Gustavson, K-H., H.B. Kison, and G. Holmgren. 1986. Prevalence of the fragile-X syndrome in mentally retarded boys in a swedish county. Amer. J. Med. Genet. **23**: 581-587.

Han, J., Hsu, C., Zhu, Z., Longshore, J.W., and Finley, W.H.: Over-representation of the disease associated (CAG) and (CGG) repeats in the human genome. Nucl. Acids Res. **22**: 1735-1740.

Harding, A.E., P.K. Thomas, M. Baraitser, P.G. Bradbury, J.A. Morgan-Hughes, and J.R. Ponsford. 1982. X-linked recessive bulbospinal neuronopothy: A report of ten cases. J. Neurol. Neurosurg. and Psychiatry **45**: 1012-1019.

Harley, H.G., J.D. Brook, S.A. Rundle, S. Crow, W. Reardon, A.J. Buckler, P.S. Harper, D.E. Housman, and D.J. Shaw. 1992. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. Nature **355**: 545-546.

Harper, D.S. and P.R. Dyken. 1972. Early-onset dystrophia myotonic: Evidence supporting a maternal environmental factor. Lancet **ii**: 53-55.

Harper, P.S. 1989. Myotonic Dystrophy, 2nd ed. W.B. Saunders Co., London.

Heitz, D., D. Devys, G. Imbert, C. Kretz, and J-L. Mandel. 1992. Inheritance of the fragile X syndrome: Size of the fragile X premutation is a major determinant of the transition to full mutation. J. Med. Genet. **29**: 794-801.

Hinds, H.L, C.T. Ashley, J.S. Sutcliffe, D.L. Nelson, S.T. Warren, D.E. Housman, and M. Schalling. 1993. Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome. Nat. Genet. **3**: 36-43.

Hoffmann, A., E. Sinn, T. Yamamoto, J. Wang, A. Roy, M. Horikoshi, and R.G. Roeder. 1990.   Highly   conserved   core   domain   and unique N terminus with presumptive   regulatory   motifs   in   a human TATA factor (TFIID). Nature **346**: 387-390.

Hoogeveen, A.T., R. Willemsen, N. Meyer, K.E. de Rooij, R.A.C. Roos, G.J.B. van Ommem, and H. Galjaard. 1993. Characterization and localization of the Huntington disease gene product. Hum. Mol. Genet. **2**: 2069-2073.

Höweler, C.J., H.F.M. Busch, J.P.M. Geraedts, M.F. Niermeijer, and A. Staal. 1989. Anticipation in myotonic dystrophy: Fact or fiction? Brain **112**: 779-797.

Huntington's Disease Collaborative Research Group. 1993. A novel gene containing a trinucletide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell **72**: 971-983

Kennedy, W.R., M. Alter, and J.H. Sung. 1968. Progressive proximal spinal and bulbar muscular atrophy of late onset: A sex linked trait. Neurol. (Minneap) **18**: 671-680.

Kimberling, W.J., M.D. Weston, C. Moller, S.L. Davenport, Y.Y. Shugart, I.A. Priluck, A. Martini, M. Milani, and R.J. Smith. 1990. Localization of usher syndrome type II to chromosome 1q. Genomics **7**: 245-249.

Knight, S.J.L., A.V. Flannery, M.C. Hirst, L. Campbell, Z. Christodoulo, S.R. Phelps, J. Pointon, H.R. Middleton-Price, A. Barnicoat, M.E. Pembrey, J. Holland, B.A. Oostra, M. Bobrow, and K.E. Davies. 1993. Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell **74**: 127-134.

Koide, R., T. Ikeuchi, O. Onodera, H. Tanaka, S. Igarashi, K. Endo, H. Takahashi, R. Kondo, A. Ishikawa, T. Hayashi, M. Saito, A. Tomoda, T. Miike, H. Naito, F. Ikuta, and S. Tsuji. 1994. Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nature Genet. **6**: 9-13.

Kouzarides, T. and E. Ziff. 1988. The role of the leucine zipper in the fos-jun interaction. Nature **336**: 646-651.

Krawczun, M.S., E.C. Jenkins, and W.T. Brown. 1985. Analysis of the fragile-X chromosome: Localization and detection of the fragile site in high resolution preparations. Hum Genet. **69**: 209-211.

Kremer, E.J., M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S.T. Warren, D. Schlessinger, G.R. Sutherland, and R.I. Richards. 1991. Mapping of DNA instability at the Fragile X to a trinucleotide repeat sequence $p(CCG)_n$. Science **252**: 1711- 1714.

Kruyer, H., M. Milà, G. Glover, P. Carbonell, F. Ballesta, and X. Estivill. 1994. Fragiel X syndrome and the $(CGG)_N$ mutation: Two families with discordant MZ twins. Am. J. Hum. Genet. **54**: 437-442.

Kuhl, P.A. and C.T. Caskey. 1993. Trinucleotide repeats and genome variation. Curr. Opin. Genet. and Devel. **3**: 404-407.

Kunkel, T.A. 1993. Slippery DNA and diseases. Nature 365: 207-208.

Ladschultz, W.H., P.F. Johnson, and S.L. McKnight. 1989. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science **243**: 1681-1688.

LaSpada, A.R., E.M. Wilson, D.B. Lubahn, A.E. Harding, and K.H. Fischbeck. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature **352**: 77-79.

Lavedan, C., H. Hofmann-Radvanyi, P. Shelbourne, J-P. Rabes, C. Duros, D. Savoy, I. Denhaupas, S. Luce, K. Johnson, and C. Junien. 1993. Myotonic dystrophy: Size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism. Am. J. Hum. Genet. **52**: 875-883.

Leach, F.S., N.C. Nicolaides, N. Papadopoulos, B. Liu, J. Jin, R. Parsons, P. Peltomäki, P. Sistonen, L. Aaltonen, M. Nyström-Lahti, X-Y. Guan, J. Zhang, P.S. Meltzer, J-W. Yu, F-T. Kao, D.J. Chen, K.M. Cerosaletti, R.E.K. Fournier, S. Todd, T. Lewis, R.J. Leach, S.L. Naylor, J. Weissenbach, J-P. Mecklin, H. Järvinen, G.M. Petersen, S. R. Hamilton, J. Green, J.Jass, P. Watson, H.T. Lynch, J.M. Trent, A. de la Chapele, K.W. Kinzler, and B. Vogelstein. 1993. Mutations of a *mutS* homolog in hereditary nonpolypsis colorectal cancer. Cell **75**: 1215-1225.

Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**: 203-221.

Lewis, R.A., B. Otterud, D. Stauffer, J.M. Lalouel, and M. Leppert. 1990. Mapping recessive opthalmic diseases: Linkage of

the locus for usher syndrome type II to a DNA marker on 1q. Genomics 7: 250-256.

Li, S-H., M.G. McInnis, R.L. Margolis, S.E. Antonarakis, C.A. Ross. 1993. Novel triplet repeat containing genes in human brain: Cloning, expression, and length polymorphisms. Genomics **16**: 572-579.

Lin, B., J. Nasir, H. MacDonald, G. Hutchinson, R.K. Graham, J.M. Rommens, and M.R. Hayden. 1994. Sequence of the murine Huntington disease gene: evidence for conservation, and polymorphism in triplet (CCG) repeat alternate splicing. Hum. Mol. Genet. **3**: 85-92.

Lindblom, A., P. Tannergård, B. Werelius, and M. Nordenskjöld. 1993. Genetic mapping of a second locus prediposing to hereditary non-polypsis colon cancer. Nat. Genet. **5**: 279-282.

Lubahn, D.B., D.R. Joseph, M. Sar, J. Tan, H.N. Higgs, R.E. Larson, F.S. French., and E.M. Wilson. 1988. The human androgen receptor: complimentary deoxyribonucleic acid cloning, sequence analysis, and gene expression in prostate. Mol. Endo. **2**: 1265-1275.

MacDonald, M.E., A. Novelletto, C. Lin, D. Tagle, G. Barnes, G. Bates, S. Taylor, B. Allitto, M. Altherr, R. Myers, H. Lehrach, F.S. Collins, J.J. Wasmuth, M. Frontali, and J.F. Gusella. 1992. The Huntington's disease candidate region exhibits many different haplotypes. Nature Genet. **1**: 99-103.

Mahadevan, M., C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barcelo, K. O'Hoy, S. LeBlond, J. Earle-MacDonald, P.J. De Jong, B. Wieringa, and R.G. Korneluk. 1992. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science **255**: 1253-1255.

Maniatis, T., J. Sambrook, and E.F. Fritsch. 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Martin, J.B. and J.F. Gusella. 1986. Huntington's disease: Pathogenesis and management. N. Engl. J. Med. **315**: 1267-1276.

Mattera, R., J. Codina, A. Crozat, V. Kidd, S.L.C. Woo, and L. Birnbaumer. 1986. Identification by molecular cloning of two forms of the α-subunit of the human liver stimulatory (G_s) regulatory component of adenylate cyclase. FEBS **206**: 36-42.

McConkie-Russel, A., A.M. Lachiewicz, G.A. Spiridigliozzi, J. Tarleton, S. Schoenwald, M.C. Phelan, P. Goonewardena, X. Ding, and W.T. Brown. 1993. Evidence that methylation of the FMR-1 locus is responsible for variable phenotypic expression of the fragile X syndrome. Am. J. Hum. Genet. **53**: 800-809.

Meier, U.T. and G. Bloebel. 1990 A nuclear localization signal binding protein in the nucleolus. J. Cell Biol. **111**: 2235-2245.

Meijer, H., E. deGraaf, D.M.L. Merckx, R.J.E. Jongbloed, C.E.M. deDie-Smulders, J.J.M. Engelen, J-P. Fryns, P.M.G. Curfs, and B.A. Oostra. 1994. A deletion of 1.6 kb proximal to the CGG repeat of the FMR1 gene causes the clinical phenotype of the fragile X syndrome. Hum. Mol. Genet. **3**: 615-620.

Meyers, R.H., M.E. MacDonald, W.J. Koroshetz, M.P. Duyao, C.M. Ambrose, S.A.M. Taylor, G. Barnes, J. Srinidhi, C.S. Lin, W.L. Whaley, A.M. Lazzarini, M. Schwarz, G. Wolff, E.D. Bird, J.P.G. Vonsattel, and J.F. Gusella. 1993. De novo expansion of $(CAG)_N$ repeat in sporadic Huntington's disease. Nature Genet. **5**: 168-173.

Mhatre, A.N., M.A. Trifiro, M. Kaufman, P. Kazemi-Esfarjani, D. Figlewicz, G. Rouleau, and L. Pinsky. 1993. Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. Nat. Genet. **5**: 184-187.

Mussbaum, R.L and D.H. Ledbetter. 1990. The fragile X syndrome. In The Metabolic Basis of Inherited Disease, ed. C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle, 327-341. McGraw Hill, New York.

Nagafuchi, S., H. Yanigisawa, K. Sato, T. Shirayama, E. Ohsaki, M. Bundo, T. Taked, T. Tadokoro, I. Kondo, N. Murayama, Y. Tanak, H. Kikushima, K. Umino, H. Kurosawa, T. Furukawa, K. Nihei, T. Inoue, A. Sano, O. Komure, M. Takahashi, T. Yoshizawa, I. Kanazawa, and M. Yamada. 1994. Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12. Nature Genet. **6**: 14-18.

Naito, H. and S. Oyanagi. 1982. Familial myoclonus epilepsy and choreoathetosis: Hereditary dentatorubral-pallidoluysian atrophy. Neurol. **32**: 798-807.

Nakahori, Y., S.J.L. Knight, J. Holland, C. Schwartz, A. Roche, J. Tarleton, S. Wong, T.J. Flint, U. Froster-Iskenius, D. Bentley, K.E. Davies, and M.C. Hirst. 1991. Molecular heterogeneity of the fragile X syndrome. Nucl. Acids Res. **19**: 4355-4359.

Neville, C.E., M.S. Mahadevan, J.M. Barceló, and R.G. Korneluk. 1994. High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. Hum. Mol. Genet. **3**: 45-51.

Ni, B., S. Rush, J.W. Gurd, and I.R. Brown. 1992. Molecular cloning of calmodulin species which are preferentially expressed in neurons in the rat brain. Mol. Brain. Res. **13**: 7-17.

Nisson, P.E., A. Rashtchian, and P.C. Watkins. 1993. Rapid and efficient cloning of Alu-PCR products using uracil DNA glycosylase. PCR Met. and Appl. **1**: 120-123.

Novelli, G., M. Gennarelli, G. Zelano, A. Pizzuti, C. Fattorini, C.T. Caskey, and B. Dallapiccola. 1993. Failure in detecting mRNA transcripts from the mutated allele in myotonic dystrophy muscle. Biochem. Mol. Biol. Int. **29**: 291-297.

O'Hoy, K., C. Tsilfidis, M.S. Mahadevan, C.E. Neville, J. Barceló, A.G.W. Hunter, R.G. Korneluk. 1993. Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. Science **259**: 809-812.

Oberlé, I., F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Boué, M.F. Bertheas, and J.M. Mandel. 1991. Instability of a 550-base pair DNA seqment and abnormal methylation in Fragile X syndrome. Science **252**: 1097-1102.

Orr, H.T. M-Y. Chung, S. Banfi, T.J. Kwiatkowski Jr., A. Servadio, A.L. Beaudet, A.E. McCall, L.A. Duvick, L.P.W. Ranum, and H.Y. Zoghbi. 1993. Expansion of an   unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature Genet. **4**: 221-226.

Oudet, C., E. Mornet, J.L. Serre, F. Thomas, S. Lentes-Zengerling, C. Kretz, C. Deluchat, I. Tejada, J. Boué, and J.L Mandel. 1993. Linkage disequilibrium between the fragile X mutation and two closely linked CA repeats suggests that fragile X chromosome are derived from a small number of founder chromsomes. Am J. Hum. Genet. **52**: 297-304.

Papadopoulos, N., N.C. Nicolaides, Y-F. Wei, S.M. Ruben, K.C. Carter, C.A. Rosen, W.A. Haseltine, R.D. Fleischmann, C.M Fraser, M.D Adams, J.C. Ventler, S.R. Hamilton, G.M. Petersen, P. Watson, H.T. Lynch, P. Peltomäki, J-P. Mecklin, A. de la Chapelle, K.W. Kinzler, and B. Vogelstein. 1994. Mutation of a mutL homolog in hereditary colon cancer. Science **263**: 1625-1629.

Peltomäki, P., L.A. Aaltonen, P. Sistonen, L. Pylkkänen, J-P. Mecklin, H. Järvinen, J.S. Green, J.R. Jass, J.L. Weber, F.S. Leach, G.M. Petersen, S.R. Hamilton, A. de la Chapelle, and B. Vogelstein. 1993. Genetic mapping of a locus prediposing to human colorectal cancer. Science **260**: 810-812.

Pieretti, M., F. Zhang, Y-H. Fu, S.T. Warren, B.A. Oostra, C.T. Caskey, and D.L. Nelson. 1991. Absence of expression of the FMR-1 gene in fragile X syndrome. Cell **66**: 817-822.

Polymeropoulos, M.H., D.S. Rath, H. Xiao, and C.R. Merril. 1991. Trinucleotide repeat polymorphism at the human transcription factor IID gene. Nucl. Acids Res. **19**: 4307.

Reik, W., E.R. Maher, P.J. Morrison, A.E. Harding, and S.A. Simpson. 1993. Age at onset in Huntington's disease and methylation at D4S95. J. Med. Genet. **30**: 185-188.

Reyniers, E., L.Vits, K. DeBoulle, B. Van Roy, D. Van Velzen, E. deGraaf, A.M.J.H. Verkerk, H.Z.J. Jorens, J.K. Darby, B. Oostra, and P.J. Williams. 1993. The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. Nat. Genet. **4**: 143-146.

Rhyner, J., M. Ottiger, R. Wicki, T. Greenwood, and E.E. Strehler. 1994. Structure of the human CALM1 calmodulin gene and identification of two CALM1-related psuedogenes CALM1P1 and CALM1P2. Submitted.

Richards, R.I. and G. Sutherland. 1994. Simple repeat DNA is not replicated simply. Nat. Genet. 6: 114-116.

Richards, R.I. and G.R. Sutherland. 1992. Dynamic mutations: A new class of mutations causing human disease. Cell **70**: 709-712.

Richards, R.I., K. Holman, K. Friend, E. Kremer, D. Hillen, A. Staples, W.T. Brown, P. Goonewardena, J. Tarleton, C. Schwartz, and G.R. Sutherland. 1992. Evidence of founder chromosomes in fragile X syndrome. Nat. Genet. **1**: 257-260.

Rousseau, F., L.J. Robb, P. Rouillard, and V.M. Der Kaloustian. 1994. No mental retardation in a man with 40% abnormal methylation at the FMR-1 locus and transmission of sperm cell mutations as premutations. Hum. Mol. Genet. **3**: 927-930.

Russel, M., S. Kidd, and M.R. Kelley. 1986. An improved filamentous helper phage for generating single-stranded plasmid DNA. Gene **45**: 333-338.

Sabouri, L.A., M.S. Mahadevan, M. Narang, D.S.C. Lee, L.C. Surh, and R.G. Korneluk. 1993. Effect of the myotonic dystrophy (DM) mutation on mRNA levels of the DM gene. Nat. Genet. **4**: 233-238.

Shaw, D.J., S. Chaudhary, S.A. Rundle, S. Crow, J.D. Brook, P.S. Harper, and H.G. Harley. 1993. A study of DNA methylation in myotonic dystrophy. J. Med. Genet. **30**: 189-192.

Sherman, S.L., N.E. Morton, P.A. Jacobs, and G. Turner. 1984. The marker (X) syndrome: A cytogenetic and genetic analysis. Ann. Hum. Genet. **48**: 21-37.

Siomi, H., M.C. Siomi, R.L. Nussbaum, and G. Dreyfuss. 1993. The protein product of the fragile X gene, FMR1, has characteristics fo an RNA binding protein. Cell **74**: 291-298.

Smits, A.P.T., J.C.F.M. Dreesen, J.G. Post, D.F.C.M. Smeets, C. de Die-Smulders, T. Spaans-van der Bijl, L.C.P. Govaerts, S.T. Warren, B.A. Oostra, and B.A. van Oost. 1993. The fragile X syndrome: no evidence for any recent mutations. J. Med. Genet. **30**: 94-96.

Soma, G-I., N. Kitahara, and T. Andoh. 1984. Molecular cloning and characterization of a cDNA clone for a protein specifically expressed in embryo as well as in a chemically induced pancreatic B cell tumor of rat. Biochem. Biophys. Res. Comm. **124**: 164-171.

Stallings, R.L. 1994. Distribution of trinucleotide microsattelites in different categories of mammalian genomic sequence: Implications for human genetic diseases. Genomics **21**: 116-121.

Strand, M., T.A Prolla, R.M. Liskay, and T.D. Petes. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature **365**: 274-276.

Strong, T.V., D.A. Tagle, J.M. Valdes, L.W. Elmer, K. Boehm, M. Swaroop, K.W. Kaatz, F.S. Collins, and R.L. Albin. 1993. Widespread expression of the human and rat Huntington's disease gene in brain and nonneural tissues. Nature Genet. **5**: 259-265.

Sturtevant, A.H. 1925. The effect of unequal crossing-over at the *Bar* locus in *Drosophila*. Genetics **10**: 117

Sutherland, G.R. and F. Hecht. 1985. Fragile Sites on Human Chromosomes. Oxford University Press, New York.

Sutherland, G.R., and E. Baker. 1992. Characterisation of a new rare fragile site easily confused with the fragile X. Hum. Mol. Genet. **1**: 111-113.

Takahashi, H., E. Ohama, H. Naito, S. Takeda, S. Nakashima, T. Makifuchi, and F. Ikuta. 1988. Hereditary dentatorubral-pallidoluysian atrophy: Clinical and pathologic variants in a family. Neurol. **38**: 1065-1070.

Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucl. Acids Res. 16: 6463-6471.

Telenius, H., B. Kremer, Y.P. Goldberg, J. Theilmann, S.E Andrew, J. Zeisler, S. Adam, C. Greenberg, E.J. Ives, L.A. Clarke, and M.R. Hayden. 1994. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. Nature Genet. **6**: 409-414.

Telenius, H., H.P.H. Kremer, J. Thielmann, S.E. Andrew, E. Almqvist, M. Anvret, C. Greenberg, J. Greenberg, G. Lucotte, F. Squiteri, E. Starr, Y.P. Goldberg, and M.R. Hayden. 1993. Molecular analysis of juvenile Huntington's disease: The major influence on $(CAG)_N$ repeat lenght is the sex of the affected parent. Hum. Mol. Genet. **2**: 1535-1540.

Thibodeau, S.N., G. Bren, and D. Schaid. 1993. Microsatellite instability in cancer of the proximal colon. Science **260**: 816-819.

van der Ven, P.F.M., G. Jansen, T.H.M.S.M. van Kuppevelt, M.B. Perryman, M. Lupa, P.W. Dunne, H.J. ter Laak, P.H.K. Jap, J.H. Veerkamp, H.F. Epstein, and B. Wieringa. 1993. Myotonic dystrophy kinase is a component of neuromuscular junctions. Hum. Mol. Genet. **2**: 1889-1894.

Verkerk, A.J.M.H., M. Pieretti, J.S. Sutcliffe, Y-H. Fu, D.P.A. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M.F. Victoria, F. Zhang, B.E. Eussen, G-J.B. van Ommen, L.A.J. Blonden, G.J. Riggins, J.L Chastain, C.B. Kunst, H. Galjaard, C.T. Caskey, D.L. Nelson., B.A. Oostra, and S.T. Warren. 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile X syndrome. Cell **65**: 905-914.

Wang, Y-H., S. Amirhaeri, S. Kang, R.D. Wells, and J.D. Griffith. 1994. Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. Science **265**: 669-671.

Warner, C.L., J.E. Griffin, J.D. Wilson, L.D. Jacobs, K. Murray, K. Fishbeck, D. Dickoff, and R.C. Griggs. 1991. X-linked spinomuscular atrophy: A kindred with associated abnormal androgen receptor binding. Neurol. **41**(Suppl. 1): 313.

Watson, D.K., M.C. Psallidopoulos, K.P. Samuel, R. Dalla-Favera, and T.S. Papas. 1983. Nucleotide sequence analysis of human c-myc locus, chicken homologue, and myelocytomatosis virus MC29 transforming gene reveals a highly conserved gene product. Proc. Natl. Acad. Sci. **80**: 3642-3645.

Wawrynczak, E.J. and R.N. Perham. 1984. Isolation and nucleotide sequence of a cDNA encoding human calmdulin. Biochem. Int. **9**: 177-185.

Weber, J.L., and P.E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. **44**: 388-396.

Wharton, K.W., B. Yedvobnick, V.G. Finnerty, and S. Artavanis-Tsakonas. 1985. *opa*: A novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *D. melanogaster*. Cell **40**: 55-62.

Whitehead, A.S., E. solomon, S. Chambers, W.F. Bodmer, S. Povey, and G. Fey. 1982. Assignment of the structural gene for the third component of human complement to chromosome 19. Proc. Nat. Acad. Sci. USA. 79: 5021-5025.

Wöhrle, D., D. Kotzot, M.C. Hirst, A. Manca, B. Korn, A. Schmidt, G. Barbi, H-D Rott, A. Poutska, K.E. Davies, and P. Steinbach. 1992. A microdeletion of less than 250 kb, including the proximal part of the FMR-1 gene and the fragile X site, in a male with the clinical phenotype of fragile-X syndrome. Am. J. Hum. Genet. **51**: 299-306.

Wöhrle, D., I. Hennig, W. Vogel, and P.S. Steinbach. 1993. Mitotic stability of fragile X in differentiated cells indicates early post-conceptional trinucleotide repeat expansion. Nat. Genet. **4**: 140-142.

Wöhrle, D., M.C. Hirst, I. Kennerknecht, K.E. Davies, and P.S. Steinbach. 1992. Genotype mosaicism in fragile X fetal tissues. Hum. Genet. **89**: 114-116.

Zoghbi, H.Y., M.S. Pollack, L.A. Lyons, R.E. Ferrel, S.P. Daiger, and A.L. Beaudet. 1988. Spinocerebellar ataxia: Variable age of onset and linkage to human leukocyte antigen in a large kindred. Ann. Neurol. **23**: 580-584.

VITA

The author, James P. Carney, was born in Youngstown, Ohio on February 22, 1967 to George and Norene Carney.

Mr. Carney attended St. Rose elementary school in Girard, Ohio and Ursuline High School in Youngstown, Ohio. In August, 1985, Mr. Carney entered Niagara University in Lewiston, New York. In May, 1989, Mr. Carney graduated from Niagara, receiving a Bachelor of Science in Chemistry/Education. In August, 1990, Mr. Carney enrolled in the Department of Molecular and Cellular Biochemistry at Loyola University Chicago, Maywood, Illinois. He joined the laboratory of Mark R. Kelley, Ph.D., in January, 1991, where he developed PCR based methodologies for the cloning of trinucleotide repeat containing cDNA fragments. In 1993, Mr. Carney received a National Institutes of Health Predoctoral Fellowship from the National Institute of Mental Health.

Mr. Carney has accepted a position as a postdoctoral fellow in the laboratory of Stuart Linn, Ph.D., at the University of California, Berkeley, California.

APPROVAL SHEET

The dissertation submitted by James P. Carney has been read and approved by the following committee:

Mark R. Kelley, Ph.D., Director
Associate Professor, Department of Pediatrics, Wells Center for Pediatric Research
Riley Hospital for Children, Indiana University Medical Center

Sally A. Amero, Ph.D.
Assistant Professor, Department of Molecular and Cellular Biochemistry
Loyola University of Chicago, Stritch School of Medicine

Michael Fasullo, Ph.D.
Assistant Professor, Department of Radiotherapy
Loyola University of Chicago, Stritch School of Medicine

John M. Lopes, Ph.D.
Assistant Professor, Department of Molecular and Cellular Biochemistry
Loyola University of Chicago, Stritch School of Medicine

Russel O. Pieper, Ph.D.
Assistant Professor, Departments of Medicine and Pharmacology, Section of Hematology/Oncology
Loyola University of Chicago, Stritch School of Medicine

The final copies have been examined by the director of the Dissertation committee and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the Dissertation is now given final approval by the committee with reference to content and form.

The Dissertation is, therefore, accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

November 29, 1994
Date

Mark R Kelley
Director's Signature