# The Molecular Basis of X-linked Agammaglobulinemia

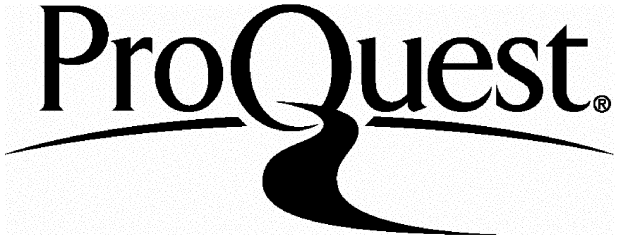
by

**Linda Anne Dugdale Bradley**

**Institute of Child Health**
**University of London**

**1995**

A thesis presented for the degree of Doctor of Philosophy



# ABSTRACT

X-linked agammaglobulinemia (XLA) is a humoral immunodeficiency disease characterised by a lack of B lymphocytes in the peripheral blood. The consequent lack of immunoglobulin results in severe infection. Affected males have pre-B cells in their bone marrow, suggesting that the defect lies in the pathway of B cell maturation.

At the start of this study, the gene responsible for XLA had not been isolated but the disease locus had been mapped to Xq22 by genetic linkage analysis. As part of a positional cloning approach to the isolation of the gene responsible for this disease, a YAC contig was constructed containing this region. A YAC was selected from this contig on the basis that it appeared not to be chimeric and that it contained the DXS178 locus, known to have no recombinations with XLA in over 70 informative meioses. This YAC was used in a cDNA enrichment study, with the aim of isolating candidate genes for XLA.

The *BTK* gene was isolated in 1993 and shown to be the gene defective in this disease. Btk is a non-receptor tyrosine kinase, related to, but distinct from Src. The identification of this gene made it possible to screen the *BTK* gene in XLA patients for mutations, using single strand conformation polymorphism (SSCP) analysis. Mutations were found and sequenced in fifteen patients, and included amino acid substitutions, small insertions and deletions, premature stop codons, disruption of the initiation codon and splice site recognition sequence alterations.

A study of the *BTK* gene in carrier women showed that in some families carrier detection could be improved by direct mutation analysis. The detection of an polymorphism within *BTK* may allow carrier detection in families where the pathological mutation has not been identified. A preliminary study of the Btk protein suggested that it will be possible to correlate the disease-causing mutation with its effect on the protein, further characterising the molecular basis of XLA in these families.

## ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr Christine Kinnon, for all her support and encouragement during the last three years. I would also like to thank Dr Ruth Lovering for all her help and advice during this study, Dr Tracy Lester for her invaluable help and patience during the writing of this thesis and Dr Steve Hinshelwood for his many words of wisdom. I would also especially like to thank everybody in the lab for being great people to work with and for ensuring that lab life was entertaining. I am indebted to numerous patients and clinicians for blood samples for this study, particularly Dr Adam Finn, Dr Mike Tarlow, Dr Kevin Forsyth, Mrs Shelia Cochrane and Dr David Goldblatt. I would especially like to thank Dr Alison Jones for all of her help. I would also like to thank the following people; Lesley Alterman for her help over the last three years, Dr Marie-Anne O'Reilly for her advice in initially starting the YAC study, Dr Paul Baird for advice on SSCP analysis, Dr David Vetrie, Dr Tony Monaco and the HGMP Resource Centre for providing YACs for this study, Dr Don Williamson and Dr Peter Jones for providing yeast size markers, Dr Michel de Weers for providing the BLCL276 cell line, Joanna Davies, Dr David Vetrie and Dr P. de Jong for providing primer sequences, Dr Paschalis Sideras for providing the *BTK* genomic sequences prior to publication and to Dr Angela Sweatman, Dr Fay Katz, Dr Bobby Gaspar and Sally Genet for discussing their unpublished observations. I really appreciated the support from my family, especially my parents, Peter and Jane, and I want to thank all my friends and ever tolerant flatmates just for being there, especially Claire, Ruth, Irené, Laurella and Caroline.

# TABLE OF CONTENTS

## CHAPTER 2

## MATERIALS AND METHODS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREIVIATIONS

| | |
|---|---|
| βARK | β adrenergic receptor kinase |
| θ | recombination fraction |
| 5-Mc | 5' methyl cytosine |
| A | alanine |
| aa | amino acid |
| ALL | acute lymphocytic leukaemia |
| ARAM | antigen recognition activation motif |
| Arg | arginine |
| ARS | autonomous replicating sequences |
| *ATK* | agammaglobulinemia tyrosine kinase |
| ATP | adenosine triphosphate |
| BCR | B cell receptor |
| BHK | baby hamster kidney |
| bp | base pair(s) |
| *BPK* | B cell progenitor tyrosine kinase |
| BSA | bovine serum albumin |
| *btk* | Bruton's tyrosine kinase gene (murine) |
| Btk | Bruton's tyrosine kinase protein |
| *BTK* | Bruton's tyrosine kinase gene (human) |
| cDNA | complementary DNA |
| cen | centromere |
| CHEF | contour clamped homogeneous electric field |
| cM | centiMorgan |
| CML | chronic myeloid leukaemia |
| CMM | chemical mismatch analysis |
| D | aspartic acid |
| dATP | deoxyadenosine triphosphate |
| dCTP | deoxycytidine triphosphate |
| ddF | dideoxyfingerprinting |
| DEPC | diethylpyrocarbonate |
| DGGE | denaturing gradient gel electrophoresis |
| dGTP | deoxyguanosine triphosphate |
| DMSO | dimethylsulphoxide |
| dNTP | deoxynucleoside triphosphate |
| dpm | disintegrations per minute |
| DTT | dithiothreitol |
| dTTP | deoxythymidine triphosphate |
| DW | distilled water |
| E | glutamic acid |
| EBV | Epstein Barr Virus |
| EDTA | ethylene diamine tetra acetic acid (disodium salt) |
| F | phenylalanine |
| $Fc_\epsilon RI$ | IgE Fc receptor I |
| FCS | foetal calf sreum |
| FDC | follicular dendritic cells |

| | |
|---|---|
| FISH | flourescent *in-situ* hybridisation |
| *g* | gravity |
| G | glycine |
| GAP | GTPase activating protein |
| *GLA* | α galactosidase A gene (human) |
| Gly | glycine |
| GST | glutathione-S-transferase |
| H | histidine |
| HIGM1 | X-linked hyper IgM syndrome |
| HLA | human leukocyte antigen |
| HS | human sonicated (DNA) |
| ICRF | Imperial Cancer Research Fund |
| IGHD | isolated growth hormone deficiency |
| IL2 | interleukin 2 |
| Itk | IL2 inducible tyrosine kinase |
| IPTG | isopropylthiogalactoside |
| K | lysine |
| kb | kilobase |
| LB | Luria-Bertani medium |
| LINE | long interspersed nuclear elements |
| LOD | logarithm of odds |
| *ltk* | leukocyte tyrosine kinase |
| M | methionine |
| M-MLV | Moloney-Murine leukaemia virus |
| Mb | megabase |
| mRNA | messenger RNA |
| N | asparagine |
| NMR | nuclear magnetic resonance |
| OD | optical density |
| OLB | oligonucleotide labelling buffer |
| P | proline |
| PBS | phosphate buffered saline |
| PBS-T | PBS tween |
| PCR | polymerase chain reaction |
| PFGE | pulsed field gel electrophoresis |
| pfu | plaque forming units |
| PH | pleckstrin homology |
| PIP2 | phosphatidylinositol biphosphate |
| PKC | protein kinase C |
| PLCγ2 | phospholipase Cγ2 |
| PLP | proteolipid protein |
| PMSF | phenyl methyl sulphonyl fluoride |
| PTK | protein tyrosine kinases |
| Q | glutamine |
| R | arginine |
| RFLP | restriction fragment length polymorphism |
| RNA | ribonucleic acid |
| RT-PCR | reverse transciptase PCR |

| | |
|---|---|
| S | serine |
| *S.cerevisae* | *Saccharomyces cerevisae* |
| SDS | sodium dodecyl sulphate |
| SDS-PAGE | SDS polyacrylamide gel electrophoresis |
| SH | Src homology |
| SINE | short interspersed nuclear elements |
| snRNA | small nuclear RNA |
| SSC | saline sodium citrate |
| SSCP | single strand conformation polymorphism |
| SSPE | sodium chloride/sodium phosphate/EDTA |
| T | threonine |
| TAE | Tris/acetate/EDTA |
| TBE | Tris/borate/EDTA |
| TE | Tris/EDTA |
| tel | telomere |
| tRNA | transfer RNA |
| Trp | tryptophan |
| TTF | Tris/tween/foetal calf serum |
| Tyr | tyrosine |
| U | unit |
| UV | ultraviolet |
| V | valine |
| W | tryptophan |
| WAS | Wiskott-Aldrich syndrome |
| WASP | Wiskott-Aldrich syndrome protein |
| X-CGD | X-linked chronic granulomatous disease |
| X-Gal | 5-bromo-4-chloro-3-indolyl-$\beta$-D-galactoside |
| X-SCID | X-linked severe combined immunodeficiency |
| *xid* | X-linked immunodeficiency (murine) |
| XLA | X-linked agammglobulinemia |
| XLP | X-linked lymphoproliferative disease |
| Y | tyrosine |
| YAC | yeast artificial chromosome |
| YEPD | yeast extract/peptone/D-glucose medium |
| Z | maximum LOD score |

# CHAPTER 1
# INTRODUCTION

## 1.1 The isolation of disease-related genes

### 1.1.1 The molecular basis of inherited disease

An understanding of the molecular basis of an inherited disease requires a full knowledge of both the gene responsible and the protein it encodes. The DNA sequence, pattern of inheritance and mechanisms of control of expression of the gene, the spectrum of pathological mutations and the effect of protein function and dysfunction on the cellular physiology and biochemistry must all be understood. A knowledge of the molecular basis of genetic disease can then lead to further understanding of cellular structure and/or protein function and can have clinical benefits in the diagnosis of disease and carrier detection. The initial step in the path to the understanding of the molecular basis of inherited disease is the isolation of the disease-related gene.

### 1.1.2 Approaches to the identification of genes

The most recent estimate of the number of genes in the human genome is 60,000 - 70,000 (Fields *et al.* 1994). The identification of a particular disease-related gene from the $3x10^9$ base pairs of the human haploid genome poses many technical difficulties. There are three main approaches which can be taken in order to identify an unknown disease-related gene; functional cloning, the candidate gene approach and positional cloning.

### 1.1.3 Functional cloning

Functional cloning can be used when the protein product involved in a particular disease has been previously identified, and knowledge of the protein is used to aid the identification of its encoding gene. For example, antibodies can be used to enrich for

specific mRNA by polysome immunopurification, as shown by Russell *et al.* (1983). The authors of this study isolated the bovine low density lipoprotein receptor gene, which underlies familial hypercholesterolemia in humans, from the enriched mRNA by screening with oligonucleotides whose sequence was based on knowledge of a small amount of amino acid sequence. Sakai *et al.* (1994) isolated the human galactocerebrosidase gene, responsible for Krabbe disease, using oligonucleotides designed in the same way. Thus, the previous identification of a protein can aid in the isolation of the associated gene.

### 1.1.4 The candidate gene approach

As the number of human genes which have been isolated increases, and the methods for identifying mutations become more sophisticated, the candidate gene approach is being more widely used. Previously identified genes can be considered as disease locus candidates if the encoded protein can been envisaged to have a role in the physiology or biochemistry of the diseased tissue, or if the gene is known to map to the same chromosomal region as the disease and encode a plausible protein. For example, many genes involved in phototransduction have already been isolated and one of these, the rhodopsin gene, was found to be responsible for dominant retinitis pigmentosa on the basis of the identification of mutations in the gene in patients (Dryja *et al.* 1990).

### 1.1.5 Positional cloning

In many cases, however, the protein associated with a particular disease pathology is not known and cannot be guessed at, so the gene must be isolated in another way. Positional cloning involves the identification of a gene on the basis of its location in the genome. Cytogenetic analysis can provide valuable information if any affected individuals have chromosomal rearrangements or cytogenetically visible deletions. These genetic alterations can be visualised and used to provide "boundaries" of the potential location of the disease gene. The identification of the genes for Duchenne muscular dystrophy (Ray *et al.* 1985), von Recklinghausen neurofibromatosis (Fountain *et al.* 1989) and choroideremia (Cremers *et al.* 1990) all involved characterisation of large deletions found

in the DNA of affected individuals. Without chromosomal deletions or translocations in patient DNA, a frequently used approach to the isolation of disease genes is saturation genetic and physical mapping. The identification of the approximate position of the gene by genetic linkage analysis of affected families, followed by physical mapping of the area to determine the order of loci and the distance between them, can allow the identification of a region, in the order of a few megabases of DNA, in which the gene must lie. Small deletions, which are not cytogenetically detectable, in affected individuals may be detected by Southern blot analysis using probes from the region, and may further narrow down the area. In order to finally locate and isolate the gene in question, cloned DNA from the critical area can be used to isolate candidate genes. Positional cloning has been successfully applied in the identification of many disease-related genes, including those responsible for cystic fibrosis (Rommens *et al.* 1989) and X-linked chronic granulomatous disease (Royer-Pokora *et al.* 1986).

### 1.1.6 Genetic linkage analysis

The initial step in many positional cloning attempts is to localise the gene in question to a particular part of a chromosome using genetic linkage analysis. This type of analysis requires polymorphic markers, such as restriction fragment length polymorphisms or microsatellite repeats, and pedigrees containing heterozygous individuals. The basis of genetic linkage analysis is that two loci which are located close together on a chromosome will be less likely to have a recombination occurring between them at meiosis I, and therefore are more likely to be inherited together, than two loci which are far apart on the same chromosome or on different chromosomes. Each chromosome undergoes at least one recombination at each meiosis. The unit of genetic map distance, the Morgan, is defined as the length of chromosomal segment which, on average, undergoes one exchange during meiosis per individual chromatid strand. Over short chromosomal regions, the recombination fraction, *i.e.* the proportion of recombinants against all possible opportunities for recombination ($\theta$), is directly proportional to the genetic map distance, so that a recombination fraction of 0.01 corresponds to a genetic map distance of 1 centiMorgan (cM). Loci which are more than 50cM apart on the same chromosome, having a recombination fraction of 0.5, will have the same chance of being

inherited together as loci on different chromosomes. As human families are generally small, an indirect statistical approach must be taken to assess genetic linkage, and this is based on maximum likelihood estimates. A ratio is established between the likelihood that the two loci are linked and the likelihood that they are unlinked. This ratio is expressed as a logarithm (base 10) - the LOD score or Logarithm of Odds - and is usually calculated using special computer programmes (for example, Lathrop & Laloulel, 1988). Evidence for linkage is considered significant when the maximum LOD score (Z) is greater than 3 (i.e. the likelihood ratio exceeds 1000). Usually, the linkage data from several pedigrees are summed and the recombination fraction is taken as that at which the LOD score is maximum. Often, multi-locus linkage analysis is used to place an unknown locus (often the disease) on a linkage map within a framework of markers, or to order a series of linked loci. Once a genetic disease locus has been assigned to a particular part of a chromosome, the area can be physically mapped as the next step towards the location of the gene. The human genome has been estimated to have a genetic size of 33 Morgans and a physical length of $3 \times 10^9$ base pairs. Very approximately 1cM therefore equals 1Mb, but this relationship varies depending on the recombigenicity of the region under study (Hartley *et al.* 1984).

### 1.1.7 Physical mapping

Physical mapping is the identification of a physical linkage between two loci by finding common restriction fragments, thus determining the order and physical distance between DNA loci on the chromosome. This procedure was revolutionised by the development of pulsed field gel electrophoresis (Schwartz & Cantor, 1984; reviewed in O'Reilly & Kinnon, 1990), which overcomes the pore-size limitations of conventional electrophoresis and allows the separation of DNA fragments of up to 10Mb in length. The original technique employs electrophoresis of DNA in alternate directions in approximately perpendicular electrical fields. When the field direction is changed after the set pulse time, shorter DNA fragments will reorientate themselves and begin to move forwards much more quickly than longer fragments. The resolution can be controlled by altering the pulse time between changes in the direction of the electrical field - higher molecular weight DNA fragments will be separated using longer pulse times. A modification of the

original method uses contour-clamped homogeneous electric fields (CHEF) (Chu *et al.* 1986) in which the hexagonal array of electrodes, at 120°, results in the DNA running straight in the tracks, allowing better resolution. The generation of fragments of DNA up to a megabase in size for pulsed field gel analysis requires careful DNA preparation to prevent shearing, usually by embedding the cells in agarose blocks before lysis. High molecular weight DNA, prepared in these blocks, is then digested into large fragments using rare-cutting restriction enzymes, singly and in double digests. Southern blot analysis of pulsed field gels with sequential hybridisation of probes to the membrane can identify common restriction fragments, which can then provide a maximum physical distance between two loci. Physical mapping can therefore allow the physical distance between loci surrounding a disease locus to be established and provide the framework within which cloned DNA can be isolated and analysed.

### 1.1.8 CpG islands

The CpG dinucleotide is under represented in the majority of the human genome because the cytosine is often methylated and this can become deaminated to give thymine (Ehrlich & Wang, 1981). CpG islands are short (about 1.5kb) regions of the genome which are generally unmethylated, and therefore C and G rich (Bird, 1987), and which seem to be clustered in the early-replicating and gene rich Giemsa light staining bands of chromosomes (Craig & Bickmore, 1994). These islands are present at the 5' ends of most mammalian housekeeping genes and approximately 40% of genes with a tissue specific or limited expression (Larsen *et al.* 1992), and to date there is no published example of a CpG island which does not appear to be associated with a gene (Bickmore & Bird, 1992). The exact function of these islands remains to be determined but they have been associated with a more open chromatin structure which may be associated with transcription (Tazi & Bird, 1990). The rare-cutting enzymes used in long range physical mapping are usually methylation sensitive and have at least one CpG dinucleotide in their recognition sequence. Identification of sites for several rare-cutting enzymes at the same location is therefore indicative of a CpG island, and so of transcribed sequences. The isolation of these islands can lead to the identification of the associated genes (Lindsay & Bird, 1987; Manoni *et al.* 1991; Tribioli *et al.* 1994).

### 1.1.9 Cloning genomic DNA

Small fragments of DNA can be cloned into plasmids (up to 10kb), bacteriophage lambda (up to 15kb) or cosmids (up to 40kb) using standard procedures (Sambrook *et al.* 1989). For positional cloning purposes however, the sizes of fragments of cloned DNA generally need to be much larger, in the order of megabases. The most commonly used vectors are yeast artificial chromosomes (YACs) which can contain over a megabase of foreign DNA (Burke *et al.* 1987). The artificial chromosome technology has bridged the resolution gap between conventional genetic analysis with resolution to 5cM (approximately 5Mb) and the DNA fragments which can be cloned in cosmids (up to 50kb). The technology involves the cloning of foreign DNA into a vector containing sequences which allow stable maintenance of the cloned DNA as an extra chromosome in the appropriate vector host cell, and its replication along with the host cellular DNA.

### 1.1.10 Generation of YACS

YACs are vectors containing up to 1-2 Mb of cloned DNA which can be propagated and maintained in yeast as an extra chromosome (Burke *et al.* 1987). The basic structure of a YAC is illustrated diagrammatically in Figure 1.1. The vector used most frequently as the basis for the YAC is pYAC4. The YAC vector sequences contain the cis-acting elements essential for yeast chromosome function: telomeres (TEL), origins of DNA replication (autonomous replicating sequences, ARS) and centromeres (CEN). These sequences confer structural integrity on the cloned DNA and allow it to replicate and have meiotic and mitotic stability. Digestion of the pYAC4 vector with *Bam*HI and *Eco*RI results in three fragments - the two larger YAC arms (*Eco*RI-*Bam*HI) and a smaller stuffer fragment (*Bam*HI-*Bam*HI). The YAC vector arms have been arbitrarily defined as left, containing the ARS and CEN4 sequences and with *trp1* as a selectable marker, and right, containing *ura3* as a selectable marker. The cloned DNA is inserted into *Eco*RI cut ends of the two vector arms, and the *Bam*HI cut ends are recognised by *Saccharomyces cerevisae* and converted *in vivo* into functional telomeres (Murray *et al.* 1986). The foreign DNA is usually size fractionated and inserted into the *Eco*RI cloning site within the *sup4* gene, before being transformed into yeast spheroplasts. *Sup4* is a

mutant tRNA which suppresses the ochre (nonsense) mutation in the *ade2* gene. Interruption of this gene by cloned DNA disrupts suppression and, in an *ade2* ochre host, will result in accumulation of a red adenine pre-metabolite, allowing selection or recombinant clones by their red colour instead of the wild type white.

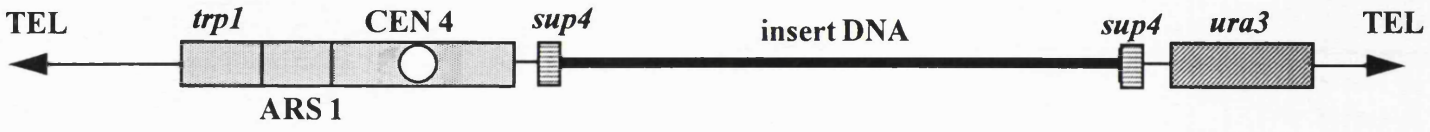
### 1.1.11 Isolation of specific YACs

A number of whole genome (Brownstein *et al.* 1989; Anand *et al.* 1990; Albertsen *et al.* 1990; Larin *et al.* 1991; Chumakov *et al.* 1992) and chromosome specific (Wada *et al.* 1990; Lee *et al.* 1992; Sleister *et al.* 1992) YAC libraries have now been generated and these have reported average insert sizes from 275kb to 918kb. These libraries can be screened using PCR amplification of a specific sequence or by hybridisation of a specific probe, to identify YACs from the area of the genome under study. YAC technology has been amenable to automation in the processes of generating filters for hybridisation (Bentley *et al.* 1992) or pools of DNA for screening by the polymerase chain reaction (PCR) (Green & Olson, 1990).

### 1.1.12 Identification of overlapping YAC clones forming a contig

YACs identified from libraries by hybridisation or PCR amplification, using loci from the region under study, can be built up into a contiguous array (contig) by identifying overlapping clones. Contigs of up to 8Mb have now been generated covering sections of the human genome (for example, Little *et al.* 1992; Monaco *et al.* 1992). YACs can be studied by PFGE and Southern blot analysis and, if there are enough known genes or DNA markers in the region, the contig can be constructed simply by using the presence or absence of these loci in the clones. The presence of a common locus in the region of interest in two or more YACs suggests overlapping cloned DNA. More usually, there are not enough known markers to allow this approach in isolation and, consequently, new probes have to be generated. An approach which has been used successfully is the generation of probes by inter-*Alu* PCR (Cole *et al.* 1991). *Alu* sequences are repeats which are thought to occur approximately every 4kb in human DNA (Moyzis *et al.* 1989). The sequences are often inverted and therefore, a single primer with sequence

**Figure 1.1. Diagrammatic representation of a YAC**




The left vector arm contains the sequences for *trp1* selectable marker gene, autonomous replicating sequences (ARS) and the *CEN4* centromeric sequences. The right arm contains the *ura3* selectable marker gene. The insert DNA is coned into the *sup4* gene. The TEL sequences function as telomeres in the yeast cell. Single thin lines represent pBR322 sequences, boxes represent yeast derived sequences and the thick single line represents the insert DNA.

based on one end of the *Alu* sequence will often amplify the intervening sequence between two repeats in opposite configurations. When applied to genomic DNA, *Alu* PCR generates a smear of product due to the number of sites to which the primer can anneal, but when used on a limited source of human DNA, for example a YAC or a somatic cell hybrid, a number of discrete fragments are produced. These fragments are characteristic for each clone as a type of "fingerprint", and can be gel purified and used as probes on previously isolated YACs or in a library screen to isolate more YACs in a chromosome walking approach. *Alu* PCR fingerprinting can also be used to establish the degree of overlap of two YACs - the bigger the proportion of identically sized bands, the greater the amount of overlap of the inserts (Nelson *et al.* 1989). New probes can also be generated by isolation of the YAC insert DNA ends using the techniques of plasmid end rescue (Burke *et al.* 1987), *Alu*-vector PCR (Nelson *et al.* 1989), inverse PCR (Silverman *et al.* 1989) or vectorette PCR (Riley *et al.* 1990) and the newly generated probes can then be used to identify further overlaps, orientate YACs relative to each other, or as a basis for screening for new YACs.

### 1.1.13 Mapping of the DNA within a YAC

Restriction digest mapping of a YAC by partial digestion, using restriction enzymes and subsequent PFGE and Southern blot analysis, allows distances between probes already known to be present in the YAC to be established and a restriction map of insert DNA to be generated. DNA loci can also be mapped relative to the ends of the YAC. This type of analysis allows the cloned DNA to be placed on the physical map of the region. Yeast DNA is not methylated however, and as many of the rare-cutting restriction enzymes are methylation sensitive, restriction digest patterns seen in DNA which has been propagated in yeast may not be easily comparable with patterns seen in genomic DNA from a cell line.

### 1.1.14 Fidelity of representation of genomic DNA in YACs

YACs can be chimeric molecules, containing DNA from two different genomic locations, or have deletions or other rearrangements, resulting in the map of the cloned DNA being

incompatible with known physical maps. The available YAC libraries have reported chimeric YAC frequencies of 10-50% (Kouprina *et al.* 1994, and references therein), but the accuracy of these figures depends on the density of study of particular libraries, and the frequency appears to be region specific. Some chromosome specific YAC libraries are reported to have lower frequency of chimeras than the whole genomic libraries (Sleister *et al.* 1992). Chimerism in YACs is thought to occur because of coligation of two genomic fragments during the preparation of the library, or because of transformation induced recombination with a co-transformed YAC (Albertsen *et al.* 1990; Green *et al.* 1991; Kouprina *et al.* 1994). The junction of one chimeric YAC has been fully analyzed and found not to have an *Eco*RI site, indicative of coligation of fragments during the cloning procedure, but to contain an *Alu* repeat sequence (Green *et al.* 1991). This suggests that recombination between two *Alu* sequences occurred after transformation. YACs have also been reported to contain internal deletions (Neil *et al.* 1990; Albertsen *et al.* 1990) which are thought to be caused by transformation induced recombination, prompted by nicks in the DNA. There have been reports of genomic regions thought to be unclonable in YACs (Palmieri *et al.* 1993). Any study involving YACs must include steps to establish the integrity of the cloned DNA, usually by fluorescent in-situ hybridisation (FISH) (reviewed in Trask, 1991) or by making insert end probes which can then be mapped onto somatic cell hybrid DNA containing the region under study (Cox *et al.* 1990). It is very important to ensure that the cloned DNA is compatible with the physical map of the region under study as failure to recognise chimera or rearrangements can lead to the misdirection of chromosome walks or wrong estimations of physical distance between two loci.

### 1.1.15 Identification of transcribed sequences from YACs

Once the area known to contain the disease gene loci has been isolated in cloned form, the next step in the gene hunting process is to identify the protein coding sequences. The identification of transcribed sequences by the conventional methods of searching for cross-species conservation (Monaco *et al.* 1986) or isolating CpG islands (Lindsay & Bird, 1987; Tribioli *et al.* 1994) cannot be feasibly undertaken when the genomic region under study spans several megabases. Even the conceptually elegant techniques of exon

trapping (Duyk *et al.* 1990; Buckler *et al.* 1991), which uses functional splicing signals to rescue transcriptional units, or enhancer trapping (Weber *et al.* 1984), which relies on the enhanced transcription of a reporter gene to identify enhancer sequences, have only been used on genomic DNA isolated as cosmids (less than 40kb). YACs, with their much larger inserts, have yet to be shown to be amenable to this type of analysis. Direct screening of cDNA libraries using radio-labelled YACs has been used successfully (Wallace *et al.* 1990; Elvin *et al.* 1990) but has low sensitivity because of the complexity of the probe. In contrast, cDNA enrichment techniques have been shown to be quick, sensitive and amenable to use with multiple samples. The hybridisation of cDNA libraries to YAC DNA immobilised on nylon membrane has been used successfully to enrich for cDNAs encoded in a particular genomic region and has lead to the isolation of particular genes (Parimoo *et al.* 1991; Lovett *et al.* 1991). The latest development of this technique involves the biotinylation of the YAC DNA and its capture using streptavidin coated magnetic beads, thereby allowing cDNA hybridisation and washing procedures to be carried out in solution (Morgan *et al.* 1992; Korn *et al.* 1992). cDNA screening, however, relies on the availability of a suitable cDNA library, the selection of which must be based on a knowledge of the temporal and developmental expression of the gene in question.

### 1.1.16 X-inactivation studies

For X-linked diseases, X-inactivation studies can provide information on the cell types in which expression of a functional gene is essential for the survival of the cell, and hence the cell types in which the gene must be expressed. In order to achieve dosage equivalence between XX females and XY males, one X chromosome is transcriptionally silenced in every cell of the early developing female embryo (reviewed in Rastan, 1994). This occurs in an entirely random manner and is stable and heritable in subsequent cell divisions. If one of the X chromosomes is carrying a mutation in a gene required for the development of a particular cell type however, all of the cells of that type which inactivate the X chromosome carrying the normal copy of the gene will not survive. The remaining cells of that particular type will have a non-random pattern of X-inactivation. Analysis of X-inactivation patterns can therefore identify the cell type and developmental

stage where the gene exerts its effect, and at which it must therefore be expressed, allowing selection of a suitable cDNA library. The techniques which are used to determine X-inactivation patterns are often based on the difference in methylation status of the active and inactive X chromosomes and make use of methylation sensitive restriction enzymes to study the pattern of inactivation (Vogelstein *et al.* 1987; Boyd & Fraser, 1990).

### 1.1.17 Confirming the identity of a candidate gene

Once a cDNA mapping to the defined area of the genome has been identified, it can be considered as a candidate gene for the disease in question. It must then be established that the gene is expressed in the cell type(s) affected by the disease, by northern blotting, RT-PCR or *in situ* hybridisation, and that the gene is defective in individuals affected by the disease. The quickest way to initially screen for genomic alterations in a particular gene is by Southern blot analysis of patient DNA, using the cDNA as a probe, to identify missing or altered fragments. Intragenic deletions or mutations which can be proposed to be disease causing firmly associate the gene with the disease. In many cases, however, no alterations can be detected using this technique and point mutations must be found using a screening technique or directly by DNA sequence analysis.

### 1.1.18 Analysis of the protein product involved in a disease phenotype

The end point of a positional cloning study is the isolation of the gene which, when defective, causes the disease under study. Whole new areas of investigation are opened up to study the function of the encoded protein in normal cells and to elucidate the molecular basis of the disease. Database searches can identify homologous proteins, imparting clues to the function of the protein and studies can be initiated to examine the expression, function, regulation and interactions of the protein. This can lead to the biochemical pathways and molecular interactions of the protein in the cell being more fully understood. A knowledge of the molecular basis of the disease may then lead to improved, and more specific treatments for affected individuals, including the possibility of treatment by gene therapy for the disease, in the future.

## 1.2 Mutation Analysis

### 1.2.1 The importance of mutation analysis

The reasons for identifying pathological mutations in a disease related gene are three-fold: to show that it is a disease causing gene; to provide functional information about the encoded protein; and to allow better disease diagnosis and provide the means for improved genetic counselling of families affected by the disease, through the use of improved carrier detection and pre-natal diagnosis.

### 1.2.2 Mutation analysis techniques

Cytogenetic and Southern blot analyses are often used as an initial screen to identify large deletions or other rearrangements in a gene. The majority of disease-causing mutations, however, are not detectable using these techniques. The unequivocal way to search for unknown point mutations in a gene is to sequence the entire coding region, splicing signal sequences and control regions. Multiple samples and large genes make this approach impractical, although automated DNA sequencing technologies may soon make it a more feasible option. Screening methods offer a more efficient way of identifying mutations from large numbers of patients. These techniques usually involve screening the gene in fragments and identifying one fragment which may contain the mutation, thereby limiting the DNA which must be sequenced. The method chosen depends on the size and structure of the gene, the sensitivity required and the resources available. There are several techniques currently being used successfully to detect single base changes, small deletions and insertions.

### 1.2.3 Single strand conformation polymorphism analysis (SSCP)

SSCP analysis (Orita *et al.* 1989a; Orita *et al.* 1989b) has been shown to identify up to 95% of mutations in a given gene (Michaud *et al.* 1992; Sheffield *et al.* 1993) and has been used widely because of its sensitivity and simplicity (for example, Hayward *et al.* 1994; Shimizu *et al.* 1994). The technique works on the principle that single stranded

DNA fragments analysed on a non-denaturing polyacrylamide gel form secondary structures which are sequence dependent. A single base alteration in a given fragment can lead to the formation of slightly different secondary structure, causing the fragment to migrate differently under electrophoresis, compared to a normal fragment. Radiolabelling of the fragments allows the change to be identified as a band shift on an autoradiograph of the polyacrylamide gel. Once the fragment in which the mutation lies has been identified, sequencing of the DNA in that fragment will locate the exact mutation. A variation on the original technique is RNA SSCP, which has been reported to detect mutations in cases where DNA SSCP did not (Sarkar *et al.* 1992b; Danenberg *et al.* 1992). Non-radioactive visualisation methods such as ethidium bromide staining (Hongyo *et al.* 1993) and silver staining (Oto *et al.* 1993) have been used widely to increase the simplicity and safety of the technique. Fluorescence based detection of shifted fragments using automated DNA sequencing technology (Iwahana *et al.* 1994) may provide the future direction for the technique of SSCP.

### 1.2.4 Denaturing gradient gel electrophoresis (DGGE)

Several other methods for the detection of point mutations are based on the identification of heteroduplexes formed between a mutant and a normal allele. These heteroduplexes can be analyzed by DGGE (Myers *et al.* 1985; Sheffield *et al.* 1989). The gradient of denaturing agent (urea and formamide) causes the DNA to denature as it migrates, in sequence dependent domains of low melting temperature. Usually, one end of the fragment is "clamped" by using a long GC tailed PCR primer in the amplification reaction, to ensure that melting occurs from one end of the molecule only. Once the two strands are partially denatured, migration through the gel is impeded. Mutations are detected by observing a DNA fragment with a different melting temperature than normal, visualised as a shift on the polyacrylamide gel. DNA sequencing of the region contained in the shifted fragment is required to identify the nucleotide change. Setting up the optimum denaturing conditions for each fragment is time-consuming but, once established, DGGE provides a mutation detection rate of up to 100% (Moyret *et al.* 1994). The technique has been modified to use temperature instead of chemical denaturants, temperature gradient gel electrophoresis (TGGE), and has been found to

have equivalent detection rates (Scholz *et al.* 1993).

### 1.2.5 Heteroduplex analysis

Heteroduplexes of normal and mutant DNA analysed on non-denaturing gels can be used to identify point mutations (White *et al.* 1992), as heteroduplexes are thought to move more slowly than corresponding homoduplexes due to a more "open" double stranded DNA configuration surrounding the mismatched bases. New gel matrices, for example Hydrolink gels, have been reported to improve sensitivity (Keen *et al.* 1991). The fragment in which the change is detected must be sequenced to identify the exact nucleotide alteration. Heteroduplex analysis has also been shown to be useful in carrier detection (Prior *et al.* 1994).

### 1.2.6 Chemical mismatch analysis (CMM)

CMM has been acclaimed as the most sensitive of the mutation detection methods currently available (Cotton *et al.* 1988). This technique involves chemically modifying mismatched bases in a heteroduplex containing radiolabelled wild type DNA, and then cleaving with piperidine at the modified bases. Osmium tetroxide is used for the modification of mispaired thymines and hydroxylamine for mispaired cytosines, with adenine and guanine mismatches being detected by labelling the anti-sense strand of the wild-type DNA. The fragment will be broken at the site of the mutation. Analysis on a polyacrylamide gel allows sizing of the cleaved fragments and therefore close estimation of the position of the altered base(s). Sequencing then reveals the nature of the mutation. This technique has been shown to pick up 100% of mutations in fragments up to 1.7kb (Forrest *et al.* 1991; Roberts *et al.* 1992). The methodology however, is labour intensive and involves using highly toxic chemicals as well as radioactivity.

### 1.2.7 Dideoxy fingerprinting (ddF)

The technique of ddF (Sarkar *et al.* 1992a) is a hybrid between dideoxy sequencing and SSCP analysis. A ladder of bands is generated using one of the four standard dideoxy

sequencing reactions and then subjected to non-denaturing polyacrylamide gel electrophoresis. Changes in the banding pattern compared to a normal control indicate the approximate position of the mutation. This technique has not been widely used but has been shown to detect 100% of mutations in a given gene (Sarkar *et al.* 1992a).

### 1.2.8 Mechanisms of mutagenesis

Changes to the DNA content of a cell occur frequently and can range from changes in the number of chromosomes to a single base substitution. Changes which do not involve the loss or gain of genetic material, which do not occur in coding regions of the genome or do not cause any functional changes to the encoded molecule, do not usually cause disease. Other genetic changes, however, are more disruptive to the cell. The effect of these changes depends on whether the cell is a somatic cell or a germ line cell. In somatic cells, most DNA alterations do not cause disease, the principle exception being changes resulting in the deregulation of cell growth, leading to malignancy. A second exception is an alteration occurring at a very early stage of embryogenesis, resulting in a chimeric individual, who may or may not exhibit a disease phenotype. The majority of disease-causing mutations occur in the germ-line and cause disease in succeeding generations.

#### 1.2.8.1 Chromosomal alterations

Non-disjunction during meiosis can result in an abnormality in the number of chromosomes in a zygote. Many alterations of this type are lethal but some changes are viable, for example Down's syndrome (reviewed in Serra & Neri, 1990). Translocations, the interchange of chromosome fragments, can result in defects and disease if part of the genetic material is lost or if, in a balanced translocation, the breakpoint occurs in or near a protein coding sequence (Schinzel, 1984). Chromosomal alterations of this type are cytogenetically visible.

### 1.2.8.2 Large insertions and deletions

Most genetic disease, however, is caused by genetic alterations which are not cytogenetically visible. Inversions of 500kb within the Factor VIII gene have been shown to be responsible for 50% of severe cases of Haemophilia A (Lakich *et al.* 1993). This is thought to be due to recombination between a small gene found in the intron of the Haemophilia A gene and a copy of this gene found 500kb towards the telomere (Rossiter *et al.* 1994). Some genes have been shown to be highly prone to deletions, for example the genes encoding steroid sulphatase (Ballabio *et al.* 1989), dystrophin (Niemann Seyde *et al.* 1992) and low density lipoprotein receptor (Langlois *et al.* 1988).

### 1.2.8.3 Trinucleotide expansions

Trinucleotide repeat expansions are a recently discovered mechanism which can cause genetic disease. Diseases including Huntingdon's disease, fragile X mental retardation and myotonic dystrophy have been shown to be caused by the rapid expansion of the numbers of repeats in the particular gene associated with the disease (reviewed in Bates & Lehrach, 1994). These diseases are often characterised by anticipation, *i.e.* increasing severity of the disease as it is passed on through the generations, and there is often a sex bias in disease transmission. The mechanism producing this genetic instability remains unknown.

### 1.2.8.4 Small insertions and deletions

Small deletions and insertions (less than 20 base pairs) have been found to cause a wide variety of genetically inherited diseases (Cooper & Krawczak, 1991; Krawczak & Cooper, 1991) and are thought to be caused mainly by the local sequence environment. Direct and inverted repeats, monotonic runs of bases and other sequences which cause pauses or disruptions to the action of the DNA polymerases during DNA replication are thought to be responsible for the majority of these sequence alterations.

### 1.2.8.5 Single base alterations

Single base alterations can be caused by the deamination of methylated cytosine in a CpG dinucleotide to thymine. This has been shown to be a frequent mechanism of point mutation (Cooper & Youssoufian, 1988). Environmental agents, for example ionising radiation, UV light or chemicals, can result in mutations if the DNA sequence is not faithfully repaired (reviewed in Weeda *et al.* 1993).

### 1.2.8.6 Errors in DNA replication

Spontaneous errors during DNA replication or repair account for some base alterations. DNA replication maintains faithful duplication of the genome through the specific demands of the DNA polymerases for the precise geometry of the Watson-Crick base pair. It has been estimated that the error rate of DNA polymerases is approximately $10^{-6}$ (Lewin, 1993). This rate is improved by the proof-reading activities of the DNA polymerases, which use 3'→ 5' exonuclease activity to excise any base which does not fit exactly into the growing DNA chain. Therefore, the mutation frequency has been estimated at $10^{-8}$ to $10^{-10}$. Most of the published work on DNA polymerase fidelity and editing has been done in prokaryotic systems, but it is thought that similar mechanisms exist in eukaryotic cells (Echols & Goodman, 1991).

## 1.2.9 Implications of mutation analysis

Patients affected with a disease provide natural models of protein dysfunction with obvious phenotypic effects. The study of these mutations and their correlation with any measurable biochemical, physiological or cellular effects and the patient disease state can provide information on the molecular basis of the disease. The correlation between defects in a single gene and an inherited disease may also lead to the improved diagnosis, treatment and counselling for families affected by the disease. Before the identification of the gene responsible for a particular disease, the techniques for carrier detection or pre-natal diagnosis are limited and can usually only give a risk estimate for a particular case. Genetic linkage analysis with closely linked, informative polymorphic markers can

provide a good estimate of risk, if enough family members are available. Once the gene responsible for a particular disease has been isolated, however, the possibilities for accurate carrier detection and for pre-natal diagnosis increase. The use of RFLPs within a gene allow a much more accurate risk estimate to be given because there is almost no risk of a recombination having occurred between the polymorphic site and the disease locus, except in very large genes. Once the precise mutation causing the disease in the family under study has been identified, unequivocal detection of the mutant allele can be performed using altered restriction fragment patterns or DNA sequencing, allowing simple carrier detection and prenatal diagnosis.

## 1.3 X-linked agammaglobulinemia

### 1.3.1 The seven X-linked immunodeficiency diseases

There are at least seven genes involved in immunodeficiency diseases mapping to the X chromosome, and these are summarised in Table 1.1. The genes responsible for all of these diseases, except one, have now been identified, and illustrate the successful use of both positional cloning and the candidate gene approach to the isolation of human disease genes. Genes can be preliminarily mapped to the X chromosome on the basis of an X-linked pedigree, with males affected by the disease and females transmitting the disease as carriers. This has resulted in extensive study of the X chromosome. Gene cloning attempts are also often aided by the identification of X-inactivation patterns in the cells of carrier women, as this can result in the identification of the expression pattern of the gene. X-inactivation has been widely used in the study of the X-linked immunodeficiencies as the affected cell types are found in the peripheral blood, making samples easily available.

The candidate gene approach lead to the identification of the gene for X-linked severe combined immunodeficiency (X-SCID) (reviewed in Leonard *et al.* 1994). This disease, characterised by a complete lack of T cells and abnormal B cell function, had been genetically mapped to Xq11-Xq13 and X-inactivation analysis of carriers had suggested

Table 1.1 The X-linked immunodeficiency diseases

| Disease | Cell types which have non-random X-inactivation pattern in carrier women | Chromosomal location | Gene | Reference |
|---|---|---|---|---|
| X-linked chronic granulomatous disease (X-CGD) | none | Xp21 | subunit of cytochrome $b_{558}$ | Royer-Pokora *et al.* 1986 |
| X-linked severe combined immunodeficiency (X-SCID) | T cells<br>B cells* | Xq13 | IL2 receptor $\gamma$ chain | Puck *et al.* 1993<br>Noguchi *et al.* 1993 |
| X-linked hyper IgM syndrome (HIGM1) | none | Xq26 | CD40 ligand | Allen *et al.* 1993<br>Aruffo *et al.* 1993<br>Korthauer *et al.* 1993<br>DiSanto *et al.* 1993 |
| Wiskott-Aldrich syndrome (WAS) | T cells<br>B cells<br>granulocytes<br>monocytes<br>platelets | Xp11 | WASP transcription factor? | Derry *et al.* 1994 |
| Properdin deficiency | none | Xp11.23 - Xp21.2 | properdin | Nolan *et al.* 1991<br>Nolan *et al.* 1992 |
| X-linked lymphoproliferative syndrome (XLP) | none | Xq25 | ? | Skare *et al.* 1993 |
| X-linked agammaglobulinemia (XLA) | B cells | Xq22 | Bruton's tyrosine kinase | Vetrie *et al.* 1993c<br>Tsukada *et al.* 1993 |

* in some women

that all haematopoietic lineages might express the X-SCID gene. The IL2 receptor $\gamma$ chain (Takeshita *et al.* 1992) was mapped to the Xcen-Xq13 region of the X chromosome and was suggested to be a candidate gene for this disease (Puck *et al.* 1993; Noguchi *et al.* 1993). This was confirmed by the identification of mutations in this gene in X-SCID patients (Puck *et al.* 1993; Noguchi *et al.* 1993). The IL2 receptor $\gamma$ chain is a component of the intermediate and high affinity IL2 receptors, important for IL2 signalling (reviewed in Taniguchi & Minami, 1993). The severity of the X-SCID phenotype compared to IL2 deficiency may be explained by the discovery that the IL2 receptor $\gamma$ chain is a component of multiple cytokine receptors, including the receptors for IL4, IL7 and IL15 (Kawahara *et al.* 1994).

X-linked hyper IgM syndrome (HIGM1) was mapped to the Xq26 band of the X chromosome by genetic linkage analysis (reviewed in Kroczek *et al.* 1994). This disease is characterised by selective agammaglobulinemia of the IgA and IgG subclasses. B and T cells are present at normal levels and neither type of cell shows a non-random pattern of X-inactivation in carrier women. The CD40 ligand gene was suggested to be a candidate gene for this disease on the basis that the CD40-CD40 ligand interaction was important for the maturation of the B cell response and that the gene mapped to the same region of the X chromosome as the disease. This was confirmed by the identification of mutations in the CD40 ligand gene in HIGM1 patients (Allen *et al.* 1993; Aruffo *et al.* 1993; Korthauer *et al.* 1993; DiSanto *et al.* 1993).

Properdin deficiency had been mapped to the short arm of the X chromosome by genetic linkage analysis (Goonewardena *et al.* 1988; Wadelius *et al.* 1992). The structural gene for properdin was found to map to the same locus and was immediately a candidate gene for the disorder (Goundis *et al.* 1989). The gene for properdin, a component of the alternative pathway of the human complement system, has now been cloned and characterised (Nolan *et al.* 1991; Nolan *et al.* 1992) but mutations in the gene in properdin deficient patients have yet to be identified.

The positional cloning approach was used in the identification of the genes for X-linked chronic granulomatous disease (X-CGD). The cloning of the X-CGD gene was a result

of the analysis of patients with large deletions of the Xp21 band. Selected cDNAs from the region were analysed, and one, the $\beta$ chain of the cytochrome$_{b558}$ of NADPH oxidase, was found to be mutated in X-CGD patients (Royer-Pokora *et al.* 1986). NADPH oxidase is responsible for the production of superoxide which kills micro-organisms ingested by phagocytes, defects in which result in recurrent infection, chronic inflammation and large granulomas (reviewed in Roos, 1994).

Wiskott-Aldrich syndrome (WAS) is characterised by eczema, thrombocytopenia and recurrent infections as a result of profound immunodeficiency of T and B cells. Genetic linkage analysis had mapped the disease to Xp11.22-p11.23. Direct cDNA selection using an immobilised YAC from a contig containing the critical region for the disease yeilded a cDNA from a gene which appeared to be expressed in lymphocytic and megakaryocytic lineages and was mutated in WAS patients (Derry *et al.* 1994). This gene encodes a proline rich protein, which has been suggested to be a transcription factor.

The gene responsible for X-linked lymphoproliferative syndrome (XLP) has not yet been identified. Individuals affected by XLP are unable to mount a normal antibody response to Epstein Barr Virus (EBV) infection. Exposure to EBV can lead to infectious mononucleosis, lymphoma and hypogammaglobulinemia (Grierson & Purtilo, 1987). Affected males that are not exposed to EBV infection appear completely normal and X-inactivation studies of female carriers have shown random patterns in all cell types investigated (Conley *et al.* 1990). This gene does not, therefore, appear to be required for the development of any of the haematopoietic lineages and may only be activated in response to certain types of infection. Three XLP patients have been found who all have deletions in Xq25 (Skare *et al.* 1993). This information may help in the isolation of the gene responsible for this disease.

The gene for X-linked agammaglobulinemia (XLA) has recently been identified using both positional cloning (Vetrie *et al.* 1993c) and the candidate gene approach (Tsukada *et al.* 1993), illustrating the use of both techniques, and will be the principle focus of this thesis.

### 1.3.2 XLA - the disease

Lack of immunoglobulins in a patient suffering from severe infections was described as early as 1952 by Dr Ogden Bruton, an American physician (Bruton, 1952). Now, more than forty years later, primary immunodeficiency diseases are more fully classified and understood. X-linked agammaglobulinemia (XLA; Bruton's agammaglobulinemia; MIM 30030; AGMX1) is a primary immunodeficiency disease characterised by a lack of circulating B cells and profound hypogammaglobulinemia. The principle effect of this lack of humoral immune response is serious recurrent bacterial infections, particularly in the upper and lower respiratory tract, but affected individuals are also susceptible to enteroviruses. These infections can be life-threatening at a very young age. The current treatment for this disease is aggressive antibiotic therapy and regular intravenous immunoglobulin infusions. The current estimate of the incidence of XLA is 1/150,000 (Smith *et al.* 1994b).

### 1.3.3 Normal B cell development

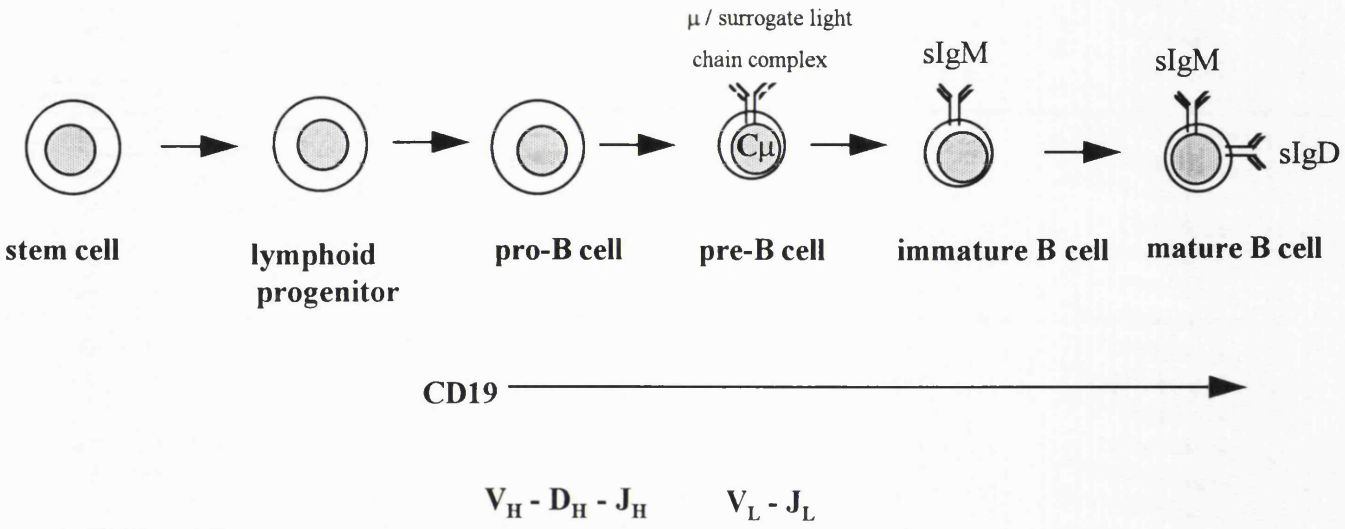
B cells are lymphocytes which carry clonally distributed surface immunoglobulin receptors for the recognition of antigen. Their development is marked by a series of developmental changes in gene expression and cell surface markers, and by immunoglobulin gene rearrangements, summarised in Figure 1.2 (reviewed in Tarlinton, 1994). The first stage of B cell development is the production of the immunocompetent primary B cell repertoire, a process which is antigen independent. In mammals, B cells are produced in the fetal liver and the adult bone marrow. Derivatives of the pluripotent haematopoietic stem cell can first be identified as having been committed to the B cell lineage when they express multiple B cell specific cell surface markers, including CD19. These are pro-B cells and have rearranged variable region genes for the $\mu$ chain, although the $\mu$ protein is not expressed at this stage. The presence of cytoplasmic $\mu$ chains is used as a marker for development into a pre-B cell and these cells divide rapidly, and some of the cytoplasmic $\mu$ is expressed on the cell surface, in association with the surrogate light chains, $\lambda 5$ and $V_{preB}$, as the pre-B cell receptor. The pre-B cell then stops dividing, becomes smaller and rearranges its light chain variable region genes. The cell then

expresses surface IgM, in association with Ig$\alpha$ and Ig$\beta$, as a functional B cell receptor (BCR), and leaves the bone marrow for the peripheral circulation as an immature B cell. Most of these cells also go on to synthesize $\delta$ chains and hence become fully immunocompetent peripheral B cells. These cells have already undergone a process of negative selection to remove any B cells expressing immunoglobulin reactive against self-antigens. Normal B cells make up 5-15% of the lymphocytes in the peripheral blood.

### 1.3.4 Normal B cell activation

In a normal healthy individual, most of the peripheral B cells are quiescent. An encounter with specific antigen is required before activation occurs and the cell can enter into cell cycle, proliferate and differentiate into mature plasma cells, which produce antigen specific antibody (reviewed in Cambier *et al.* 1994). The functions of this antibody can be direct, for example the neutralisation of bacterial toxins, or indirect, via the opsonisation of the antigen, facilitating the action of the complement cascade or phagocytosis. B cells can be activated by some types of antigen, specifically polymeric bacterial antigens, without help from other cells of the immune system but the resultant primary and secondary responses are fairly weak. The B cell response to most antigens, however, depends on T cell help in the form of cell-cell interactions and soluble factors. The B cell can then grow, proliferate and differentiate into an antibody producing cell. Other B cells migrate into follicles in the spleen and undergo class switching, to produce antibody of different isotypes, and somatic mutation in their V region genes. Antigen-antibody complexes are trapped by follicular dendritic cells (FDC) in the primary follicles and this FDC associated antigen causes selective proliferation of B cells with high affinity surface immunoglobulin. These cells migrate through the follicle and differentiate into memory cells in the germinal centres. These memory cells enter the peripheral circulation and remain in a resting state until the next encounter with antigen, when they can provide a fast response.

**Figure 1.2 Schematic representation of B cell development**




Schematic representation of B cell development showing surface immunoglobulin expression. The developmental stages at which CD19 is expressed and where immunogloubulin gene rearrangement occurs are indicated.

### 1.3.5 B cells in XLA

B cells make up less than 1% of peripheral lymphocytes in XLA patients (Conley, 1985; Campana *et al.* 1990). This failure of B cell development results in affected individuals having very small lymph nodes without germinal centres. Pearl *et al.* (1978) reported that pre-B cells with cytoplasmic $\mu$ could be detected in the bone marrow of XLA patients indicating that the defect responsible for this disease did not prevent entry of stem cells into the B cell lineage. Campana *et al.* (1990) noted that the ratio of pro-B cells to pre-B cells in the bone marrow of affected individuals was markedly increased; ten times the number of pro-B to pre-B cells were found in XLA patients, the reverse of the ratio found in normal controls. Both Campana and Milili *et al.* (1993) reported decreased pre-B cell proliferation. The few B cells found in the peripheral blood of XLA patients have an immature phenotype, as shown by more intense expression of surface IgM and less intense expression of HLA Class II (Conley, 1985; Tedder *et al.* 1985), but have been shown to have normal immunoglobulin gene recombination mechanisms and expression of the surrogate light chain genes (Anker *et al.* 1989; Timmers *et al.* 1991; Milili *et al.* 1993). The defect in this disease was shown to be intrinsic to B cells by X-inactivation studies which identified a non-random pattern of X-inactivation in the mature B cells of carrier women and a random pattern of X-inactivation in their T cells, neutrophils and granulocytes (Conley *et al.* 1986; Fearon *et al.* 1987).

### 1.3.6 Why search for the gene responsible for XLA?

The gene responsible for XLA must encode a protein which is crucial for the development of B cells, as XLA males have no peripheral mature B cells and obligate carrier women have a non-random pattern of X inactivation in their B cells. The identification of this gene should therefore provide an insight into B cell development in both the normal and diseased state. Clinically, the identification of the gene may allow better disease diagnosis and genetic counselling for families affected by this disease, and may potentially, in the future, lead to a cure for this disease by gene therapy.

### 1.3.7 Genetic mapping of the XLA locus

At the start of this study the molecular basis for XLA was unknown. Disease inheritance patterns were suggestive of an X-linked monogenic disorder. A positional cloning approach to the identification of the gene responsible for this disease was taken by several laboratories and the XLA locus was first assigned to the long arm of the X chromosome in 1986, when two studies showed linkage to DXS3 ($Z = 3.65$ at $\theta = 0.04$ (Kwan *et al.* 1986) and $Z = 3.30$ at $\theta = 0.06$ (Mensink *et al.* 1986)). Malcolm *et al.* (1987) excluded DXS3 from the XLA locus on the basis of a recombination ($Z = 3.63$ at $\theta = 0.06$) and showed close linkage between the XLA locus and DXS17 ($Z = 4.44$ at $\theta = 0$) and DXS94 ($Z = 6.65$ at $\theta = 0$). Arveiler *et al.* (1987) ordered these loci as: cen-DXS3-DXS94-DXS17-tel and showed DXS178 to be closely linked to DXS94. A recombination between the XLA locus and DXS94 identified this loci as the distal flanking marker for the disease locus (Kwan *et al.* 1990). DXS178 was shown to have no recombination with the disease in over 40 informative meioses and a combined two point LOD score of $Z = 14.48$ at $\theta = 0$ (Guioli *et al.* 1989; Kwan *et al.* 1990) indicating very close linkage. These studies lead to a genetic map of the region: cen-DXS3-6.5cM-(DXS178,XLA)-5cM-DXS94-3.5cM-DXS17-tel. New markers were then mapped to Xq21.3-Xq22 (Dietz Band *et al.* 1990; Barker *et al.* 1991), several of which (DXS265, DXS366, DXS442) were placed within the XLA region (Barker *et al.* 1991). The *Msp*I polymorphism associated with DXS101 was also placed in this region (Barker *et al.* 1991). Kwan *et al.* (1991) found no recombinations between the XLA locus and DXS265, DXS366 or DXS442 in a preliminary study, but two independent studies subsequently reported recombinations between the XLA locus and DXS366 and DXS442 (Lovering *et al.* 1993a; Parolini *et al.* 1993). As DXS366 has been placed proximal to DXS442 by a recombination (Barker *et al.* 1991), DXS442 replaced DXS3 as the proximal marker flanking XLA. Lovering *et al.* (1993a) also identified a recombination between the disease and the DXS101 associated *Msp*I polymorphism, making this the new distal flanking marker. No recombinations were detected with DXS265, which has now been shown to map within 5kb of DXS178 (Lovering *et al.* 1993b). DXS178 remained the closest marker to the XLA locus with no recombinations with the disease locus in over 70 informative meioses and the cumulative two point LOD score in excess of 30,

indicating very close linkage with the disease. The genetic distance between DXS442 and DXS101 was thought to be approximately 2-4cM, with the order of loci being cen-DXS366-DXS442-(XLA-DXS178-DXS265)-DXS101-DXS94-DXS17-tel.

### 1.3.8 Heterogeneity of the disease

There have been reports of cases of XLA where family genetic linkage studies have mapped the disease locus to a region other than Xq22 (Mensink *et al.* 1986). It appears, however, that the origin of the XLA mutation in several families, including the family described in Mensink *et al.* (1986), is in the paternal germ-line and that the father is an unaffected germ-line mosaic (Lau *et al.* 1988; Hendriks *et al.* 1989). After taking this mode of inheritance into consideration, the linkage data did not exclude an XLA locus at Xq22.

Females with a phenotype identical to XLA have been described (Conley & Sweinberg, 1992). There was no evidence for translocations involving the X chromosome in these patients, suggesting the possibility of an autosomal location for a gene resulting in an XLA-like phenotype.

A number of patients with reduced numbers of, but not absent, peripheral B cells and low immunoglobulin levels have been reported (Conley & Puck, 1988; Alterman *et al.* 1993), and in some cases there is considerable variation in severity of disease between affected individuals in the same family (Goldblum *et al.* 1974). X-inactivation studies of obligate carrier women from these families show a non-random pattern of X-inactivation in the B cell lineage (Conley & Puck, 1988; Alterman *et al.* 1993), suggesting that this mild disease may be an allelic variant of XLA.

### 1.3.9 XLA and isolated growth hormone deficiency

The co-inheritance of XLA and isolated growth hormone deficiency (IGHD) has been reported in a number of families (Fleischer *et al.* 1980; Sitz *et al.* 1990; Conley *et al.* 1991; Monafo *et al.* 1991; Buzi *et al.* 1994). Genetic linkage analysis of these families

indicated that the defect mapped to the same region of the X chromosome as XLA and the non-random pattern of X-inactivation seen in carrier women was identical to that seen in the B cells of obligate XLA carrier women (Conley *et al.* 1991). The growth hormone gene is located on chromosome 17 (Owerbach *et al.* 1980). These results suggest that XLA with growth hormone deficiency is due either to a contiguous gene deletion, resulting in loss of the XLA gene and a gene involved in growth hormone production, or to an allelic variant of the XLA gene contributing to the growth hormone defect by an unknown mechanism.

### 1.3.10 Identification of the gene responsible for XLA

In 1993, two papers, published simultaneously, announced the isolation of the gene responsible for XLA. Using a positional cloning approach, Vetrie *et al.* ( 1993c) selected cDNAs from a B cell lineage cDNA library using a YAC from a contig spanning the XLA critical region. XLA patient DNA was screened by Southern blot analysis using the cDNAs as probes, and one group of cDNA clones was found which showed altered DNA fragments in patients, but not in normal controls, indicating an association with XLA. The same gene was isolated by Tsukada *et al.* (1993), who were looking for genes encoding tyrosine kinases expressed in early B cell development. They screened a cDNA library from murine B cell progenitors with a probe encoding a kinase domain and found a cDNA which was expressed in B cells and myeloid cells, but not other cell lineages. After isolation of the human cDNA, fluorescent *in situ* hybridisation, analysis of somatic cell hybrids, and studies using a YAC contig showed that this gene mapped to the XLA critical region. Using an anti-sera directed against the protein encoded by this gene, they showed that cell lines derived from XLA patients were deficient in the protein, confirming the link between this gene and XLA.

### 1.3.11 Bruton's tyrosine kinase

The isolated human gene was initially called *ATK* (agammaglobulinemia tyrosine kinase) by one group (Vetrie *et al.* 1993c) and the mouse gene called *BPK* (B cell progenitor kinase) by another (Tsukada *et al.* 1993), but both are now known as *BTK* (Bruton's

tyrosine kinase). The *BTK* gene encodes a 2.5kb mRNA with a 1.9kb open reading frame, and shows high levels of homology with non-receptor tyrosine kinase genes of the Src family, but particularly to *ITK* (Siliciano *et al.* 1992) and *TEC* (Mano *et al.* 1993), and with these two genes, is thought to be part of a new subfamily of non-receptor protein tyrosine kinases (PTKs). From homology to these other PTKs, it can be predicted that the 659 amino acid Btk protein has the defined Src homology (SH) 1, 2 and 3 domains, thought to be involved in catalysis, protein-protein interactions involving the binding of a phosphotyrosine residue and other protein-protein interactions via a proline rich motif, respectively. The Btk/Itk/Tec subfamily differ from the other Src related PTKs in that they lack the N terminal myristoylation site in the SH4 domain, thought to be responsible for membrane association, and the C terminal regulatory tyrosine residue. They also have a much longer N terminal region, now shown to contain the recently defined Pleckstrin homology domain (Musacchio *et al.* 1993).

### 1.3.12 *xid* - a mouse model for XLA?

The mouse X-linked immunodeficiency *xid* found in the CBA/N mouse has a milder phenotype than human XLA (reviewed in Lindsberg *et al.* 1991). These mice have large numbers of peripheral B cell, but a severely reduced population of CD5 positive B cells, and almost normal levels of immunoglobulin with decreased IgM and IgG3. The main functional defect in the B cells of these mice is the complete inability to respond to some thymus-independent antigens. They also have an abnormal response to a variety of activation signals. This immunodeficiency had been postulated to be the mouse equivalent of human XLA firstly because both diseases involve the B cell lineage on the basis of X-inactivation studies. Secondly, both diseases result in cells of the B lineage being unable to mature fully, and thirdly, the mouse *xid* mutation was found to genetically map to the syntenic region of the X chromosome corresponding to the XLA locus. On this basis, the *xid* defect was considered a possible candidate for the murine equivalent of human XLA. The identification of the *BTK* gene allowed further analysis of the murine *xid* genotype. Analysis of backcross progeny showed that *xid* and *BTK* mapped together on the mouse X chromosome and subsequent sequence analysis of the *BTK* gene from CBA/N mice and from other mouse strains showed that there was an

amino acid alteration, arginine 28 changed to cysteine caused by a C to T mutation at nucleotide 219, which was found only in the immunodeficient strains (Thomas *et al.* 1993; Rawlings *et al.* 1993). This confirmed that the XLA and *xid* phenotypes are caused by mutations in the same gene. The molecular basis of the difference between the *xid* and XLA phenotypes remains to be elucidated.

## 1.4 The aims of this study

At the start of this study, the gene responsible for XLA had not been identified, although the region which must contain the gene had been narrowed down to an area of approximately 2-4cM, between DXS442 and DXS101, by genetic linkage studies. The first aim of this study was therefore to construct a YAC contig across the XLA critical region. The second aim was to use a selected YAC from this contig to screen a cDNA library for transcribed sequences from the region as candidate genes for XLA. The *BTK* gene was identified in 1993 (Vetrie *et al.* 1993c; Tsukada *et al.* 1993) and shown to be defective in individuals affected by XLA. The third aim of this project was, therefore, to identify mutations in the *BTK* gene in XLA patients, to allow important residues and domains to be defined. The fourth aim of the project was to use the information and techniques from the mutation analysis study to further analyse the *BTK* gene and to begin to study the Btk protein to analyse the role of Btk in disease, to provide functional information on the protein and to assess possibilities for improved carrier detection for this disease.

# CHAPTER 2
# MATERIALS AND METHODS

## 2.1 Reagents

All reagents were from Sigma Chemical Company Ltd. unless otherwise stated. Recipes for solutions are given in section 2.16.

## 2.2 Preparation of nucleic acid

### 2.2.1 Preparation of high molecular weight DNA from yeast

Yeast colonies from agar plates were inoculated into 10ml of media (YEPD non-selective media or AHC for YAC selection) containing 50$\mu$g/ml ampicillin and incubated at 30°C with agitation for 12-36 hours. For bulk cultures, 10ml of culture was used to innoculate an additional 500ml of the same media, which was then incubated as above. High molecular weight DNA was prepared in agarose blocks using a mini-prep method (Anand *et al.* 1989). Agarose blocks were then stored at 4°C.

### 2.2.2 Preparation of genomic DNA from blood

High molecular weight DNA in agarose blocks was prepared as described previously (O'Reilly *et al.* 1992), by Dr A. Sweatman in this laboratory. DNA in solution was prepared using standard methods (Miller *et al.* 1988) by the Clinical Genetics laboratory, Hospital for Sick Children, London.

### 2.2.3 Preparation of RNA from blood

Mononuclear cells were separated from blood collected in EDTA coated tubes (less than three days old) by Ficoll (Pharmacia) separation. In one case, where the individuals

under study were carrier women, monocytes were separated out from the mononuclear cells by allowing them to adhere to microexudate-coated surfaces (Ackerman & Douglas, 1978). RNA was then prepared from either the full mononuclear layer or the purified monocytes using acid guanidinium thiocyanate-phenol-chloroform extraction (Chomczynski & Sacchi, 1987) and stored in isopropanol at -70°C. Before use, the sample was centrifuged at 12,000*g* for 15 minutes and the RNA pellet washed in 70% ethanol.

### 2.2.4 cDNA synthesis

RNA prepared from 2 x$10^6$ mononuclear cells was resuspended in 18$\mu$l of diethyl pryocarbonate (DEPC) treated water and heated to 70°C for 5 minutes. The sample was chilled on ice and the the following were added in order; 1$\mu$l RNAguard (Pharmacia), 2.5$\mu$l 0.1M dithiothreitol (DTT) (Life Technologies), 10$\mu$l 0.1mg/ml oligo d(T) 12-18 (Pharmacia), 5$\mu$l 10mM dNTPs (Pharmacia), 2.5$\mu$l 2mg/ml BSA (NBL) and 2$\mu$l reverse transcriptase (M-MLV, 200U/$\mu$l, BRL). The cDNA was synthesised by incubation at 42°C for 90 minutes, the enzyme inactivated at 65°C for 5 minutes and the sample then stored at -20°C.

## 2.3 Quantitation of nucleic acid

### 2.3.1 Spectrophotometric quantitation

Nucleic acid was quantitated using a Philips PU8620 spectrophotometer. An adsorption of 1 $OD_{260nm}$ was taken to equal a concentration of 50$\mu$g/ml double stranded DNA, 40$\mu$g/ml single stranded DNA or RNA and 33$\mu$g/ml oligonucleotides.

### 2.3.2 Ethidium bromide estimation

The concentration of DNA could be estimated by comparing 1$\mu$l of sample with 1$\mu$l of lambda ($\lambda$) DNA (Life Technologies) dilutions (1-100$\mu$g/ml) when each was mixed with

10$\mu$l of 500ng/ml ethidium bromide and examined on a UVP utltraviolet transilluminator.

## 2.4 Polymerase chain reaction (PCR)

### 2.4.1 PCR primers

PCR primers were designed to be approximately 20 nucleotides long, with equal proportions of A+T to C+G and with a C or G at the 3' end. Primers were synthesised by Mr P. Rutland, Department of Genetics, ICH, London, or by the ICRF, London. The approximate annealing temperature of the primers was calculated according to the equation $T^{\circ} = 2n(A+T) + 4n(C+G)$ where n equals the number of residues in the primer sequence. The sequences of primers used in this study are listed in sections 2.17 and 2.18.

### 2.4.2 Reaction conditions

All PCR amplifications used approximately 50-500ng template DNA, 200$\mu$M dNTPs (Pharmacia), 1.5mM $MgCl_2$, buffer and enzyme according to the manufacturer's instructions (Promega or Bioline) and 50pmol of each primer (forward and reverse) in a final volume of 50$\mu$l, unless otherwise stated. The reaction was overlaid with mineral oil to prevent evaporation during the temperature cycling. A control reaction containing all the components except template DNA was always included to check for any contamination. Reactions were carried out by denaturing initially for 3 minutes at 94°C, followed by 30 seconds at the calculated annealing temperature, 30 seconds (for products under 750bp) or 1 minute (for larger products) at 72°C and 30 seconds denaturing at 94°C, for 30 cycles. A final annealing step was then followed by 10 minutes at 72°C to ensure that all products were elongated. Reactions were temperature cycled using a Hybaid Thermal Cycler.

### 2.4.3 Optimisation of PCR conditions

If a clean product could not be produced or no product could be amplified, the conditions

were altered by changing the annealing temperature, altering the amount of template used or performing a $MgCl_2$ titration. If a persistant co-amplification product was present or there was still no amplification after these procedures, the primer sequence composition was changed by moving it 5 or 10 bp up or downstream.

## 2.5 Restriction enzyme digestion of DNA

### 2.5.1 Complete digestion of DNA

Digestion of DNA was carried out following the enzyme manufacturer's instructions (NBL or Life Technologies) in 1 x supplied restriction enzyme buffer and an excess of enzyme (5-10 U/$\mu$g DNA).

### 2.5.2 Complete digestion of DNA in agarose blocks

Blocks were first washed three times over one hour in TE buffer (at least 5ml per block). The digestion was carried out in 0.5mg/ml BSA, 5$\mu$M spermidine and 1 x restriction enzyme buffer (supplied by manufacturer) with 30U of enzyme in a final volume of 200$\mu$l overnight at the recommended temperature.

### 2.5.3 Partial digestion of DNA in agarose blocks

For partial digestion, agarose blocks containing DNA were equilibrated first in TE buffer as above and then in 1 x restriction enzyme buffer (supplied by the manufacturer) by incubating in a volume of 500$\mu$l on ice for 30 minutes. Half blocks (5$\mu$g DNA) were placed in 100$\mu$l restriction enzyme buffer with 15U, 1U, 0.3U or 0.1U of enzyme and initially incubated on ice for 30 minutes to allow the enzyme to diffuse into the block and subsequently at 37°C for 1 hour to allow digestion. The reactions were stopped by the addition of 5$\mu$l 0.5M EDTA.

### 2.5.4 Digestion of PCR products

Unless otherwise specified, PCR products were first precipitated with 2.5 volumes ethanol and 0.1 volume 3M NaOAc, at -20°C overnight. After centrifugation at 12 000*g*, the pellet was washed in 70% ethanol and then resuspended in 1 x buffer as supplied by the manufacturer. Restriction enzyme digestion was carried out according to the manufacturer's instructions.

## 2.6 Electrophoresis of DNA

### 2.6.1 Agarose gel electrophoresis

PCR products were checked for amplification and product size on 1% agarose (International Biotechnologies Inc.) gels with 500ng/ml ethidium bromide using 1 x TAE buffer, with 1kb ladder (Life Technologies) as size markers. Gels for Southern blot analysis and preparative gels were prepared at the appropriate agarose concentration, and run using the required voltage and time for the separation of the DNA fragments under study, with either 1kb ladder (Life Technologies) or λ *Hind*III (NBL) as size markers. Samples were loaded into wells as required using 6 x DNA loading buffer. Gels were photographed on a UV transilluminator.

### 2.6.2 Pulsed field gel electrophoresis

Samples were electrophoresed on 1% agarose (SeaKem GTG, FMC) gels in 0.5 x TBE using the LKB Pharmacia Pulsaphor CHEF system and the LKB hexagonal electrode kit with appropriate run length, switch times and voltage conditions. Switch times were controlled with either the LKB Pulsaphor apparatus or a Flowgen EPS switch unit. The temperature was controlled using the LKB Multitemp II or Techne RB5 themostatic circulators to maintain 12°C. The DNA size markers used were whole yeast chromosome preparations of either *Saccharomyces cerevisiae* YP148 (provided by Dr Don Williamson, NIMR, Mill Hill, London) for DNA fragments less than 1Mb or *Hansenula wingeii*

(provided by Dr Peter Jones, MRC, Cambridge) for DNA fragments over 1Mb (Jones *et al.* 1989). During long runs the buffer was changed every three days. After electrophoresis, the gels were stained with 500ng/ml ethidium bromide in 0.5 x TBE buffer for 15 minutes, and then destained in 0.5 x TBE buffer for 15 minutes before being photographed on a UV transilluminator.

### 2.6.3 Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis for SSCP analysis, DNA sequencing and protein analysis is described in sections 2.12.2, 2.13.3 and 2.15.6, respectively. Electrophoresis for separating small DNA fragments for XLA carrier detection was carried out on 20% gels using 1.5mm spacers in 1 x TBE buffer at 50V for 16 hours. The gel was stained in 5$\mu$g/ml ethidium bromide for 20 minutes and destained for 1 hour before being photographed under UV light.

## 2.7 Isolation of DNA fragments from agarose gels

In order to isolate a DNA fragment, the sample was first electrophoresed through a 1% agarose gel or a 0.8% LMP agarose gel with 500ng/ml ethidium bromide in 1 x TAE. The required DNA fragment was excised from the gel under UV light and then purified away from the agarose using the Geneclean II (Bio 101) kit. The DNA was then quantitated using ethidium bromide fluorescence (section 2.3.2).

## 2.8 Southern blot analysis

### 2.8.1 Blotting of gels

Agarose gels for Southern blot analysis were incubated in 240mM HCl for 20 minutes to break large DNA fragments by acid depurination and then in denaturing solution for

60 minutes, changing the solution once. The gels were inverted and blotted onto Hybond $N^+$ membrane (Amersham) in denaturing solution overnight. Membranes were washed thoroughly in 2 x SSC before storage between acid-free tissue paper, or hybridisation.

### 2.8.2 Radiolabelling of probes

Probes were radiolabelled by the random priming method (Feinberg & Vogelstein, 1983; Feinberg & Vogelstein, 1984) as follows: 30-50ng of double stranded DNA in a volume of 30$\mu$l was heat denatured at 98°C for 5 minutes, mixed with 10$\mu$l OLB, 2$\mu$l BSA (10mg/ml), 50$\mu$Ci [$^{32}$P]-dCTP (ICN Flow) and 2U Klenow DNA polymerase (NBL), and incubated at room temperature for 3 hours.

### 2.8.3 Removal of unincorporated [$^{32}$P]-dCTP

A 1ml syringe was plugged with polymer wool, filled with Sephadex G50, packed by centrifugation for 3 minutes at 300*g* and the column washed with 200$\mu$l 2 x SSC. The probe labelling reaction was made up to a final volume of 200$\mu$l with 2 x SSC, loaded onto the column and centrifuged as before. The activity of the recovered probe was measured by counting 2$\mu$l using a Bioscan QC 2000 $\beta$ counter.

### 2.8.4 Preannealing of probes containing repetitive sequences

The required volume of probe to give $10^6$ dpm/ml of hybridisation solution was mixed with 1.2 volumes of 20 x SSC, 1.2 volumes of 10mg/ml human sonicated DNA and 1.4 volumes of TE. The mixture was heated to 98°C for 5 minutes before being incubated at 65°C for 1-3 hours.

### 2.8.5 Prehybridisation of membranes

Membranes were wetted in 2 x SSC before being rolled into glass hybridisation bottles (Hybaid) with mesh (Hybaid) interleaved. 10ml of hybridisation solution were added and the bottles rotated in a Hybaid oven at 65°C for at least 3 hours.

### 2.8.6 Hybridisation of membranes

Probes which did not require preannealing were denatured at 98°C for 5 minutes and added to the hybridisation solution at $10^6$ dpm/ml. Preannealed probes were added directly following the preannealing reaction. Hybridisations were incubated at 65°C for over 16 hours rotating in a Hybaid oven.

### 2.8.7 Washing of membranes after hybridisation

Membranes were washed three times in 3 x SSC/0.1% SDS at room temperature for 20 minutes and then in solutions with 2xSSC, 1xSSC or 0.5xSSC all with 0.1% SDS, dependent on the probe used (see Appendix I) at 65°C for 30 minutes. Membranes were then wrapped in plastic film (Saranwrap).

### 2.8.8 Autoradiography

Membranes were exposed to X-ray film (XAR-5, Kodak) at -70°C with two intensifying screens (Lightening Plus, Cronex, Dupont) for between 1 and 14 days. Films were developed on a Fugi RGII film processor.

### 2.8.9 Stripping membranes

Membranes were stripped of annealed probe by incubating in 2mM EDTA/1mM Tris-HCl (pH8.0)/0.1% SDS at 98°C until the solution had cooled to room temperature. The stripping procedure was checked by exposure to X-ray film as before.

## 2.9 Analysis of Yeast Artificial Chromosomes (YACs)

### 2.9.1 Selection of YACs

The YACs used in this study were 9DA6 (provided by the HGMP Resource Centre,

Harrow, Middlesex) from the ICI library (Anand *et al.* 1990), 178-1, 178-2, 178-3, 178-4, 178-5 (provided by Dr A.P.Monaco, ICRF, London) and 178-J, 178-K, 101-2, 101-8, 101-5 and 442A (provided by Dr D. Vetrie, UMDS, Guy's Hospital, London) from the ICRF library (Larin *et al.* 1991), and 178-S and 178-M (provided by Dr D. Vetrie, UMDS, Guy's Hospital, London) from the St Louis Library (Brownstein *et al.* 1989). All YACs had vector arms derived from the plasmid pYAC4 and were transformed into the host strain *S. cerevisae* AB1380.

In order to select YACs 178-J and 178-K from a YAC dot blot containing DNA from yeast containing YACs positive in a primary DXS178 screen (provided by Dr M.A.-J. O'Reilly, this laboratory), a subclone of the DXS178 specific plasmid p212/9, 212XT, was sequenced to allow specific PCR primers to be generated (Dr M.-A.J. O'Reilly, personal communication). These primers, 426F and 427R (section 2.18), were used to generate a 950bp DXS178 specific product by PCR amplifying human genomic DNA. The conditions for this reaction were standard (section 2.4.2) and used an annealing temperature of 51°C. This PCR product was purified as described in section 2.7 and used to probe the YAC dot blot using the hybridisation protocols described in section 2.8. The primers and conditions for this PCR were sent to the Human Genome Mapping Project Resource Centre, Harrow, Middlesex to screen the ICI YAC library (Anand *et al.* 1990), resulting in the isolation of YAC 9DA6.

### 2.9.2 Sizing of YACs

High molecular weight yeast DNA prepared in agarose blocks (section 2.2.1) was analysed by PFGE (section 2.6.2). After ethidium bromide staining and exposure to UV light, the YAC could be seen in some cases as an extra band in the yeast chromosomal ladder. The YAC DNA was transferred on to Hybond $N^+$ membrane by Southern blotting and the membranes were hybridised with a probe specific for DNA contatined in the YAC. The YACs could then be approximately sized by comparison with *S. cerevisae* chromosomes.

### 2.9.3 Checking the YACs using vector specific probes

Membranes containing YAC DNA were hybridised with vector specific end probes for the left and right arms of the YAC vector. These probes were generated by digesting the plasmid pBR322 to yield fragments which are present in the YAC vector pYAC4 as follows. *Escherichia coli* containing pBR322 (Sambrook *et al.* 1989) were grown in 10ml LB media containing ampicillin overnight at 37°C with agitation and plasmid was obtained using the Magic miniprep kit (Promega). A *Pvu*II/*Sal*I digest on 4$\mu$g of plasmid, using 25U of each enzyme at 37°C for 2 hours, yielded a 1.4kb fragment specific for the right arm of the YAC vector. To generate the 2.3kb *Pvu*II/ *Eco*RI fragment specific for the left arm, 16$\mu$g of pBR322 was digested first with 50U *Pvu*II at 37°C for 1 hour, and then 50U *Eco*RI and 3.8$\mu$l 1M Tris-HCl, pH8.0, was added to bring the buffer close to recommended conditions and the reaction incubated at 37°C for 2 hours. The DNA fragments were gel purified as described in section 2.7 and the fragments radiolabelled and hybridised to the membranes.

### 2.9.4 *Alu* PCR fingerprinting of YACs

*Alu* PCR fingerprinting (Nelson *et al.* 1989) was performed using 50pmol Ale1 and Ale3 PCR primers (section 2.18) in a standard PCR (section 2.4.2) with 0.2$\mu$g/$\mu$l BSA and 1.2M $\beta$-mercaptoethanol. The yeast colony containing the YAC was transferred and mixed into the PCR tube. The reaction was heated to 94°C for 4 minutes followed by 30 cycles of 94°C for 1 minute and 73°C for 3 minutes. The products were visualised on a large 2% agarose gel.

### 2.9.5 Probes used to analyse the YACs

Probes specific for loci in the region used to analyse the YACs, and the washing conditions used in Southern blot analysis are listed in Appendix I. The YACs were also analysed by PCR for the two loci DXS454 and DXS458 (Weber *et al.* 1990; Parkar *et al.* 1994) and using a CA repeat at the DXS178 locus (Allen & Belmont, 1992).

### 2.9.6 Generation of YAC end probes

End probes from the YAC inserts were generated using vectorette PCR (Riley *et al.* 1990). Vectorette libraries were generated using *Rsa*I and *Hae*III and as described except that the left hand end specific PCR primer was HYAC-C (section 2.18). End probes were purified from agarose gels (section 2.7) before being radiolabelled.

### 2.9.7 Long term storage of yeast

Overnight yeast cultures were stored in 50% glycerol at -70°C for long term storage.

## 2.10 cDNA library screening for candidate genes

### 2.10.1 Biotinylation of YAC 178-1

YAC 178-1 was separated from the other yeast chromosomes by PFGE. The band containing the YAC excised under UV light and purified using the Geneclean II kit (Bio 101). The amount of DNA recovered was estimated using ethidium bromide fluoresence (section 2.3.2). A total of 180ng of DNA was biotinylated using the Clontech Photoactivatable Biotin labelling kit according to the manufacturer's instructions.

### 2.10.2 Construction of Nalm 6 cDNA library

A Nalm 6 cDNA library was made by Dr R. Lovering, in this laboratory, using methods as described in Sambrook *et al.* ( 1989) unless otherwise stated. RNA was extracted from Nalm 6 cells with guanidinium thiocyanate followed by centrifugation in caesium chloride solutions, and poly(A)+ RNA selected using an oligo (dT)-cellulose column (Pharmacia). The first strand and second strands were synthesised and synthetic phosphorylated linkers (Pharmacia) ligated on to the ends. The cDNA was purified before being ligated to pre-digested λ gt10 (Stratagene) and packaged withe Gigapack II plus (Stratagene) according to the manufacturer's instructions. The library

($1.5x10^7$pfu/$\mu$l) was stored in SM/chloroform at 4°C.

### 2.10.3 Pre-hybridisation of YAC 178-1

The hybridisation and washing conditions were based on a previously published method (Parimoo *et al.* 1991). The biotinylated YAC DNA was prehybridised using human sonicated (HS) DNA and AB1380 yeast DNA which had been purified from high molecular weight DNA agarose blocks (section 2.7). The hybridisation mix was 5 x SSPE, 5 x Denhardts and 0.5% SDS. The pre-hybridisation conditions are shown in Table 2.1. The pre-hybridisation mixtures were overlaid with mineral oil and incubated at 65°C overnight. The oil was then removed and 1$\mu$l (samples 1 and 3) or 10$\mu$l (sample 2) of Streptavidin Dynabeads (Dynal) were added and the tubes mixed for 30 minutes at room temperature. Dynabeads with attached pre-hybridised YAC DNA were purified using the magnet (Dynal) and washed four times with 0.5ml 5 x SSPE/0.1% SDS.

**Table 2.1 Prehybridisation conditions for cDNA selection**

| Sample | YAC DNA | Yeast DNA | HS DNA | Volume of DNA | Volume Hyb. Mix | Total volume |
|---|---|---|---|---|---|---|
| 1 | 2ng | 0.8$\mu$g | - | 20$\mu$l | 20$\mu$l | 40$\mu$l |
| 2 | 10ng | 4$\mu$g | - | 200$\mu$l | 200$\mu$l | 400$\mu$l |
| 3 | 2ng | - | 2$\mu$g | 20$\mu$l | 20$\mu$l | 40$\mu$l |

### 2.10.4 Hybridisation of YAC 178-1 and the Nalm 6 cDNA library

The biotinylated YAC178-1 DNA attached to the Streptavidin Dynabeads was hybridised with 6ng/$\mu$l Nalm 6 cDNA in 5 x SSPE/5 x Denhardts/0.5% SDS in a total volume of 30$\mu$l for samples 1 and 3 and 300$\mu$l for sample 2 at 65°C, overnight. The magnetic beads with YAC and hybridised cDNA attached were then separated out using the magnet. The samples were washed in 600$\mu$l volumes with the following solutions; twice for 5 minutes in 2 x SSC/0.1% SDS at room temperature and then three times for 20 minutes in 2 x SSC/0.1% SDS, 20 minutes in 1 x SSC/0.1% SDS, 10 minutes in 0.2 x SSC/0.1% SDS

and twice for 20 minutes in 0.1 x SSC/0.1% SDS, all at 65°C. This was followed by two quick washes in 0.1 x SSC at room temperature to remove the SDS.

### 2.10.5 PCR amplification of hybridised cDNAs

The hybridised cDNAs were PCR amplified directly from the immoblised YAC DNA mixture in a final volume of 100$\mu$l using 30pmol of primers C1 and C2 (Parimoo *et al.* 1991) (section 2.18). The conditions used were standard (section 2.4.2) except that the temperature cycle used was 94°C for 1 minute, 54°C for 90 seconds and 72°C for 4 minutes. 2$\mu$l of this product was then used in a second round of amplification using nested primers A1 and A2 (Patanjali *et al.* 1991) (section 2.18) in a final volume of 100$\mu$l (5 tubes). Conditions were standard (section 2.4.2) except that the temperature cycle used was 94°C for 1 minute, 50°C for 90 seconds and 72°C for 2 minutes (Patanjali *et al.* 1991).

### 2.10.6 Subcloning of hybridised cDNAs

The ligations and transformations described here were performed in conjunction with Dr R. Lovering. The PCR amplified cDNAs were subcloned into Bluescript II KS plasmid (Stratagene) using two different approaches; ligation of *Eco*RI cut amplified cDNAs into *Eco*RI cut Bluescript, and ligation of undigested amplified cDNAs into *Sma*I digested Bluescript which had been T primed.

#### 2.10.6.1 *Eco*RI cut end ligations

The amplification products from five PCR reactions were pooled and column purified as described in section 2.8.2. The DNA was resuspended in 50$\mu$l of water and digested with *Eco*RI according to the manufacturer's instructions, phenol/chloroform purified (Sambrook *et al.* 1989), column purified (section 2.8.2) and was then diluted to give a concentration of 5ng/$\mu$l. 2$\mu$g of Bluescript KS plasmid was similarly digested with *Eco*RI and purified in the same way. The ligation reaction was set up with an approximate 3:1 molar ratio of insert ends to vector ends. The reaction contained 50ng plasmid, 30ng

cDNA, 2.5$\mu$l 10 x reaction buffer containing 1 unit of ligase (NBL) and 1.3$\mu$l of water to make the final volume up to 20$\mu$l and was incubated overnight at 4°C. As a control, 25ng of *Eco*RI digested plasmid was religated in a final volume of 10$\mu$l and incubated as above.

### 2.10.6.2 Blunt end ligations

In order to use the A overhang often left by *Taq* polymerase (Mole *et al.* 1989; Hemsley *et al.* 1989), 2$\mu$g of Bluescript vector was digested with *Sma*I according to the manufacturer's instructions and then T primed using 8U terminal transferase and 1$\mu$l of 100$\mu$M ddTTP. Incubation was carried out for 30 minutes at 37°C and then 2$\mu$l 50mM EDTA was added to stop the reaction. The DNA was then cleaned by phenol/chloroform purification (Sambrook *et al.* 1989). The ligation reaction included 50ng Bluescript, 80ng undigested cDNA, 2.5$\mu$l 10 x reaction buffer with 1 U of ligase and 15$\mu$l water to make the reaction volume up to 20$\mu$l and was incubated at 4°C overnight. 25ng of T primed cut vector was religated as a control in 10$\mu$l volume.

### 2.10.7 Competent cells for transformation

SURE cells (Stratagene) were grown overnight in LB with tetracyclin at 37°C with agitation. A 10ml innoculum of this overnight culture was then grown in 130ml of the same media until the $OD_{600nm}$ = 0.6. The culture was centrifuged for 5 minutes at 6000*g* and 2°C and the cells resuspended in ice cold 50mM $CaCl_2$ at half volume. The cells were left on ice for 1 hour before being pelleted as before and resuspended in 50mM $CaCl_2$ in 0.05 of the original volume. The cells were again left on ice for 1 hour.

### 2.10.8 Transformation of subcloned cDNAs

The products of each ligation reaction were mixed in 100$\mu$l 0.1M Tris-HCl pH7.4 and then mixed with 100$\mu$l competent SURE cells (section 2.10.7). 1ng uncut Bluescript was used as a control. The cells were heat shocked at 45°C for 2 minutes and then 1ml warm LB media was added. The cells were grown at 37°C for 20 minutes before being plated

out on LB plates with tetracyclin, ampicillin, 0.1M IPTG (Calbiochem) and 2% X-Gal and grown overnight at 37°C.

### 2.10.9 Production of colony filters

The cloning site for this vector is within the coding information for the first 146 amino acids of the β galactosidase gene (lac Z). When the vector is used in cells which carry the sequences for the carboxy terminal portion of β galactosidase, the two fragments can associate to form an enzymatically active protein, which leads to colonies producing a blue colour in the presence of the chromogenic substrate X-Gal. If a foreign DNA fragment is cloned into this cloning site, the β galactosidase fragment is non-functional and the bacterial colonies containing recombinant plasmids are white. This system allows quick, easy selection of colonies containing potentially recombinant plasmids. White colonies, which potentially contain insert, were picked and gridded out onto LB plates in duplicate with the same antibiotics, one plate containing IPTG and X-Gal and the other without. The plates were grown overnight at 37°C. These filters with grown colonies were fixed according to the protocol described by Buluwela *et al.* ( 1989).

### 2.10.10 Hybridisation of colony filters

Filters were hybridised with radioactively labelled 178-1 YAC DNA and selected cDNA as described in section 2.8. YAC DNA was labelled as described (section 2.8.2 ) and selected cDNA from sample 2 was labelled by PCR using primers A1 and A2 (section 2.18) by reducing the amount of dCTP to 1mM and adding 0.2μl [$^{32}$P]-αdCTP (ICN Flow). The temperature cycle used for this labelling reaction was 94°C for 1 minute, 50°C for 90 seconds and 72°C for two minutes. The DNA was purified using a spin column, as described in section 2.8.2.

### 2.10.11 Isolation of subcloned cDNAs

White colonies were grown in 10ml LB media and ampicillin overnight at 37°C and minipreparations of plasmid DNA made using the Magic miniprep Kit (Promega). The

isolated plasmids were digested with *Eco*RI according to the manufacturer's instructions to release the insert and analysed on 1% agarose gels to determine the insert size.

### 2.10.12 Analysis of inserts

Individual *Eco*RI digested plasmid preparations were electrophoresed on 1% LMP agarose gels and the insert fragment excised, purified and radioactively labelled as described (section 2.8.1). Each insert was hybridised to *Eco*RI digested genomic DNA, AB1380 and 178-1 DNA preparations and to EcoRI digests of all other miniprep samples, by Southern blot analysis.

## 2.11 Identification of patients for *BTK* mutation analysis

### 2.11.1 Patients used for analysis

XLA patients were males diagnosed because of recurrent infections early in childhood. All had profound hypogammaglobulinemia, with virtually no B cells but normal levels of T cells. In some families, carrier women had previously been studied using X chromosome inactivation analysis (Alterman *et al.* 1993). Patients with less severe XLA phenotypes and xid-like phenotypes were also analysed. Clinical details are given in Appendix II.

### 2.11.2 Measurement of B cell numbers and immunoglobulin levels

Patient B cell numbers and immunoglobulin levels were from hospital records where possible. In the cases where B cell numbers were not available, B cells were measured using fluorocytometric analysis on a FACScan (Becton Dickinson) by staining Ficoll purified mononuclear cells with phycoerythrin (PE) conjugated mouse anti-Leu-12 (CD19) (Becton Dickinson) with PE conjugated mouse IgG1 (Becton Dickinson) as a control antibody.

## 2.12 Single strand conformation polymorphism (SSCP) analysis

### 2.12.1 PCR conditions

For SSCP analysis, the *BTK* cDNA was PCR amplified in seven overlapping sections and digested into small fragments as detailed in Table 2.2. The PCR was performed using 200$\mu$M dGTP, dATP and dTTP and 20$\mu$M dCTP (Pharmacia) with 1$\mu$Ci [$^{32}$P]-dCTP (ICN Flow), 1.5mM $MgCl_2$, 50 pmol each primer (forward and reverse), buffer and enzyme according to the manufacturer's instructions (Bioline) and 1$\mu$l cDNA (section 2.2.4) in a final volume of 25$\mu$l. Initial denaturation was at 94°C for 3 minutes and the subsequent 30 temperature cycles were at 94°C, 57°C and 72°C for 30 seconds each. Primer sequences are listed in section 2.17. The PCR products were digested when necessary by the addition of 10 units of enzyme to the PCR mix and incubation at 37°C overnight. The amplification products were checked on 1% agarose gels and then diluted between two-fold and twenty-fold in 0.1% SDS/10mM EDTA as required to allow even loading on the polyacrylamide gels.

### 2.12.2 Non-denaturing polyacrylamide gels

2$\mu$l of diluted sample and 2$\mu$l of loading buffer were denatured at 95°C for 3 minutes, chilled on ice and loaded onto a 6% polyacrylamide (Accugel, National Diagnostics) non-denaturing gel. Samples were run at 4°C on gels containing 10% glycerol in 1 x TBE, at 60 Watts for 6 hours and at room temperature on gels containing 5% glycerol in either 1 x TBE or 0.5 x TBE, at 30 Watts for 4 hours. Gels were dried and exposed to autoradiographic film (Kodak XAR-5) at room temperature for 3-14 days. Any differences in mobility between patient and control was scored as positive.

**Table 2.2 PCR primers and restriction digest sites for SSCP analysis**

| Reaction | Primers[1] | Position in cDNA[2] | Product Size (bp) | Enzyme | Cut Site[3] | Fragment sizes (bp) |
|---|---|---|---|---|---|---|
| A | A11F<br>E3R | 1829-2132 | 304 | Rsa I | 2023 | 195<br>119 |
| B | BS5(F)<br>11R | 1587-1870 | 284 | *Hae* III | 1754 | 168<br>116 |
| C | AS5(F)<br>10R | 1294-1690 | 397 | *Eco* RI | 1453 | 160<br>237 |
| D | 849F<br>7R | 964-1336 | 373 | *Ava* II | 1107 | 144<br>229 |
| E | NTSH3F<br>5R | 619-1019 | 401 | *Hae* III | 793 | 175<br>226 |
| F | A1F<br>2R | 319-664 | 346 | *Hae* III | 459 | 141<br>205 |
| G | Start F<br>Ex1R | 76-379 | 304 | | | 304 |

[1] F and R indicate forward and reverse primers, respectively. Primer sequences are given in section 2.17.
[2] Nucleotide position as numbered in Vetrie *et al.* ( 1993c).
[3] 5' cut site of enzyme.

## 2.13 DNA sequencing

### 2.13.1 Sequencing using the "Sequenase" kit version 2.0 (USB)

Template DNA for sequencing was produced by using PCR amplification of cDNA with one of each pair of primers biotinylated using standard PCR conditions (section 2.4.2) with an annealing temperature of 57°C. Singled stranded template was separated using M-280 Streptavidin Dynabeads (Dynal) according to the manufacturer's instructions. Sequencing was carried out using the "Sequenase" kit version 2.0 as detailed in USB protocols.

### 2.13.2 Cycle sequencing

For cycle sequencing, PCR product from the amplification of cDNA in a standard 50$\mu$l reaction was excised from a 1% agarose gel and purified by centrifuging for 2 minutes through polymer wool at 12,000*g*. DNA was precipitated from the resulting solution by the addition of 0.1 volumes of 3M NaOAc and 2.5 volumes EtOH at -20°C overnight. The DNA was pelleted at 12,000*g* for 15 minutes, washed in 70% EtOH, air dried and resuspended in water. Cycle sequencing was carried out according to the manufacturer's instructions using Taq Cycle-sequencing kit (USB) or Exo (-) Pfu Cyclist DNA sequencing kit (Stratagene).

### 2.13.3 Sequencing gels

Sequencing reactions were run on 6% polyacrylamide (Accugel 40, National Diagnostics) 8M urea denaturing gels in 1 x TBE, using wedged spacers and BRL sequencing equipment according to the manufacturers' instructions. Gels were dried on a vaccuum drier and exposed to autoradiographic film (Kodak XAR-5) for 1 - 14 days at room temperature.

## 2.14 Tissue culture

### 2.14.1 Culture of cell lines

Daudi is a Burkitt's lymphoma B cell line. BLCL-276 was derived from bone marrow from an XLA patient by EBV transformation and was provided by Dr M. de Weers (Leiden, The Netherlands). These cell lines were grown in RPMI 1640 (Life Technologies) supplemented with 10% FCS (Globepharm), 2mM L-glutamine (Life Technologies) and 50$\mu$l/ml gentamicin (Roussel) at 37°C under 5% (v/v) $CO_2$. Baby hamster kidney (BHK) cells were grown in the same media until confluent and then detached from the flask by incubation with 10mM EDTA in PBS (Ackerman & Douglas, 1978). AnLy and RAG are fibroblast lines and were grown in the same media, with

Trypsin-EDTA (Gibco) used to loosen the cells. AnLy is a somatic cell hybrid containing Xq12-Xqter as its only human DNA. RAG is the murine parent cell line.

### 2.14.2 Long term storage of cells

Cells were frozen at $2x10^7$/ml in growth media with 30% FCS and 10% DMSO by first cooling slowly to -70°C in a polystyrene box for 48 hours and then transferring to liquid nitrogen for long term storage. To thaw, cells were warmed quickly at 37°C, pelleted by centrifugation and resuspended in the appropriate medium.

### 2.14.3 Counting of cells

Cells were counted using a Neubauer haemocytometer according to the manufactuer's instructions.

## 2.15 Analysis of Btk protein

### 2.15.1 Production of anti-Btk antibody

An antiserum specific for Btk was made by Dr R. Lovering and Dr S. Hinshelwood, in this laboratory (Genevier *et al.* 1994b). The *BTK* cDNA sequence encoding amino acids 163 - 218 (Vetrie *et al.* 1993c) was amplified by PCR using the primers NTSH3F and SH3NTR (section 2.17) and was cloned, using the BamHI and *Eco*RI enzyme sites encoded in the PCR primers, into the bacterial expression plasmid pGEX-2T (Smith & Johnson, 1988). The resulting glutathione-S-transferase (GST) Btk fusion protein was expressed and was purified using glutathione beads (Smith & Johnson, 1988). The fusion protein could be seen by SDS-PAGE as one major species of approximately 34kD with an additional degradation product of 29kD. Approximately 50$\mu$g of this protein was used to immunize New Zealand white rabbits according to a published protocol (Tsukada *et al.* 1993) except that the rabbits were boosted monthly. The animals were bled at monthly intervals, starting one week after the first boost and serum collected and stored

at -70°C. Immunoglobulin was purified by binding to protein A Sepharose. The polyclonal antiserum was purified further by removing species which bound to either agarose, GST-sepharose or to a T cell acetone extract (Sambrook *et al.* 1989). Following western blot analysis of B cell lysates, the resulting anti-sera strongly recognised a protein of approximately 77kD (Btk).

### 2.15.2 Hypotonic lysis of cells for Btk analysis

Mononuclear cells or monocytes, purified as described in section 2.2.3, were lysed in hypotonic lysis buffer at a concentration of $5x10^7$/ml on ice for 10 minutes. The samples were centrifuged at 12,000*g* for 5 minutes at 4°C. The supernatant was removed and added to an equal volume of 2 x reducing loading buffer.

### 2.15.3 Immunoprecipitation of Btk

Btk was immunoprecipitated from $5x10^6$ mononuclear cells or purified monocytes. Where these cell numbers were not available, volumes were adjusted accordingly. Cells were washed in PBS 'A' (Oxoid, Unipath) and left on ice for 10 minutes in 100$\mu$l immunoprecipitation lysis buffer. Cell debris and unlysed cells were removed by centrifugation at 12,000*g* for 10 minutes at 4°C. The supernatant was transferred to a fresh tube containing 100$\mu$l 10% protein A Sepharose/normal rabbit serum which had previously been prepared as detailed in section 2.15.4. The sample was incubated at 4°C for 30 minutes, the protein A Sepharose was pelleted at 12,000*g* for 30 seconds and the precleared supernatant transferred to a tube containing 1.5$\mu$g of anti-Btk antibody. The sample was incubated at 4°C overnight with rotation.

To immunoprecipitate the protein, further incubation was then carried out at 4°C for 30 minutes with the addition of 20$\mu$l of 10% protein A Sepharose prepared as in 2.15.4. The Sepharose beads with the attached immunoprecipitates were then washed three times with 1ml of lysis buffer, pelleting as before, and transferring to a fresh tube following each wash. An equal volume of 2 x SDS protein sample reducing loading buffer was then added if the sample was to be analysed on a SDS polyacrylamide gel or the sample

was washed further to carry out a kinase assay (section 2.15.5).

### 2.15.4 Preparation of protein A Sepharose for immunoprecipitations

For preclearing, the required amount of 50% protein A Sepharose was incubated with a half volume of normal rabbit serum in 20mM Tris-HCl, pH8.0/130mM NaCl/1% NP40/1mM PMSF and rolled at room temperature for 30 minutes. The beads were pelleted at 12,000*g* for 30 seconds, washed three times in 0.5ml lysis buffer and resuspended at 10% in lysis buffer. For the immunoprecipitation step, the beads were simply washed three times in lysis buffer as above.

### 2.15.5 Kinase assay

In order to perform a kinase assay, the immunoprecipitated samples were washed further once in PBS 'A', twice in 0.5M LiCl/20mM Tris-HCl, pH8.0, and once in kinase buffer. The pellet was resuspended in 30$\mu$l kinase buffer with 5-10$\mu$Ci [$^{32}$P]-$\gamma$ATP (Amersham International), incubated at room temperature for 10-20 minutes and an equal volume of 2 x SDS protein sample reducing loading buffer added. The samples were then analysed by SDS PAGE.

### 2.15.6 Western blot analysis

Proteins were separated on 7.5% SDS polyacrylamide denaturing gels overnight at 40V and room temperature using Bio-Rad II gel apparatus according to the manufacturer's instructions. The electrophoresed proteins were transferred on to Hybond C nylon membranes (Amersham International) using a semi-dry blotter (Bio-Rad) according to the manufacturer's instructions using transfer buffer. The membrane was first blocked with 5% non-milk fat (Marvel)/PBS for 3 hours and then with 3% BSA/PBS for 1 hour, washing briefly in between with 0.05% Tween/PBS (PBS-T). The membranes were washed again in PBS-T before being incubated with a 1/200 dilution of anti-Btk antiserum in TTF with 0.02% thiomersal at room temperature for 3 hours or overnight. After washing 5 times over 30 minutes in PBS-T, the membrane was incubated in horse-radish

peroxidase-conjugated goat anti-rabbit Ig anti-serum diluted 1/1000 in TTF for 90 minutes. Following a further five washes in PBS-T, the proteins were detected using the ECL system (Amersham International) with film exposure times from 30 seconds to two hours.

## 2.16 Solutions

Distilled water (DW) was used to prepare all solutions and solutions were autoclaved at 121°C for 20 minutes to sterilise, unless otherwise stated.

AHC selective YAC media

1.7g yeast nitrogen base without amino acids and without ammonium sulphate (Difco), 5.0g ammonium sulphate, 10g casein hydrolysate-acid (low salt), 20g glucose, water to 1 litre, pH to 5.8. 2ml adenine hemisulphate 10mg/ml added after autoclaving. 20g/litre bacto-agar (Difco) was added for AHC agar plates.

Ampicillin stock

50mg/ml in water, filter sterilised, working concentration 50$\mu$g/ml.

Denaturing solution

1.5M NaCl, 0.5M NaOH.

100 x Denhardts solution

20g Ficoll 400 (Pharmacia), 20g polyvinylpyrrol-idone, 20g BSA (Fraction V), water to 1 litre. Sterilised by filtering and stored at -20°C.

DEPC water

DEPC added to 0.1% (v/v), incubated overnight at room temperature and autoclaved.

6 x DNA loading buffer

1.5g Ficoll in 10ml water, bromophenol blue, xylene cyanol.

Hybridisation solution

10 x Denhardts, 50$\mu$g/ml sonicated salmon sperm DNA (denatured at 98°C for 5 minutes), 6 x SSC and 1% SDS.

Hypotonic lysis buffer

10mM Tris-HCl, pH7.4, 2mM $MgCl_2$, 1mM $CaCl_2$, 2mM PMSF (BDH).

Immunoprecipitation lysis buffer

20mM Tris-HCl, pH8.0, 130mM NaCl, 1% NP40 (BDH), 1mM PMSF, 10mM NaF (BDH), 1% aprotinin, 20$\mu$M leupeptin, 1mM dithiothreitol, 100$\mu$M sodium orthovanadate

IPTG

2g IPTG in 10ml water, filter sterilised.

Kinase buffer

20mM $MnCl_2$ (BDH), 50mM Tris-HCl, pH7.5

LB (Luria-Bertani) bacterial growth medium

10g Bactotryptone (Difco), 5g bacto yeast extract (Difco), 10g NaCl, water to 1 litre. 15g/litre bacto-agar (Difco) added for LB agar plates.

OLB was made from the following solutions

Solution O : 1.25M Tris-HCl, 0.125M $MgCl_2$, pH8.0 (store at 4°C )

Solution A : 1ml Solution O, 18$\mu$l 2-mercaptoethanol and 5$\mu$l each of dATP, dTTP and dGTP previously dissolved at 0.1M in TE (stored at -20°C)

Solution B : 2M HEPES, pH 6.6 (stored at 4°C)

Solution C : Hexanucleotides (Pharmacia) suspended in TE at 90 OD units/ml

Solutions A,B and C were mixed in a ratio of 10:25:15 and stored at -20°C.

PBS-T

0.05% Tween in PBS'A' (previously autoclaved).

2 x SDS protein sample reducing loading buffer

The following were added to 40ml of DW: 1.52g Tris base, 20ml glycerol, 2g SDS, 2ml 2-mercaptoethanol, 1mg bromophenol blue (BDH). 1M HCl was added until the pH was 6.8 and the solution made up to 100ml with DW.

Sephadex G50

10g Sephadex G50 (Pharmacia) was incubated in 100ml 2 x SSC at 65°C for 3 hours, replacing the supernatant with new SSC during the incubation. Stored at 4°C.

SSC, 20 x stock

3M NaCl, 0.3M sodium citrate.

SSCP/Sequencing loading buffer

95% Formamide, 20mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol (BDH).

1 x SSPE 0.15M NaCl, 0.02M sodium phosphate, 1mM EDTA, pH7.2

Tetracyclin stock

5mg/ml in ethanol, not sterilised, stored at -20°C, diluted 1/100 for use.

Tris acetate buffer (TAE), 50 x stock

0.2M Tris base, 1M glacial acetic acid (BDH), 50mM EDTA, pH8.0.

Tris borate buffer (TBE), 10 x stock

0.9M Tris base, 0.9M Boric acid, 1mM EDTA, pH8.0.

Tris-EDTA (TE), pH8.0

10mM Tris-HCl, pH8.0, 1mM EDTA.

TTF

200mM Tris-HCl, pH7.5, 0.5% Tween, 10% FCS (Globepharm).

Western blot transfer buffer

48mM Tris, 39mM glycine, 20% methanol (BDH), pH9.2.

X-Gal

20mg/ml in dimethylformamide, store -20°C, no need to sterilise.

YEPD yeast growth media

10g bacto yeast extract, 20g peptone (Difco), 20g D-glucose, water to 1 litre, pH to 5.8. 20g/litre bacto-agar (Difco) was added for agar plates.

## 2.17 *BTK* gene primers

### Table 2.3 Sequence of primers used for PCR and sequencing

| Name | Sequence | 5'position[1] | Comments |
|---|---|---|---|
| StartF | AGCTACCTGCATTAAGTCAGG | 76 | Bio[1] |
| A1F | TGTGTTGAAACAGTGGTTCC | 319 | |
| 3F | AGAAGAGGTGAAGAGTCCAG | 373 | |
| NTSH3F | AATTCGGATCCATGGGCTGCCAAATTTTGGAG | 619 | BamHI[2] |
| 849F | ATGTATGAGTGGTATTCCAAACAC | 964 | |
| EX6F | CGTCATTATGTTGTGTGTTCCAC | 1126 | |
| AS5(F) | GCAGGCCTGGGATACGGATC | 1294 | Vetrie *et al.* 1993c |
| BS5(F) | CCTGAGGGAGATGCGCCAC | 1587 | Vetrie *et al.* 1993c |
| A11F | CGGAAGTCCTGATGTATAGC | 1829 | |
| EX1R | CTCTTCTCGGAATCTGTCTTTC | 379 | |
| 4R | TTTGAGCTGGTGAATCCACC | 519 | |
| 2R | CAGGTTTTAAGCTTCCATTC | 664 | |
| SH3NTR | AATCAGAATTCCTTTTTCAGCTCACTTGTGG | 787 | *Eco*RI[2] |
| 5R | TTTAGCAGTTGCTCAGCCTG | 1019 | Bio[1] |
| 847R | GAGACACTGGATATTTGAGCCTGG | 1267 | |
| 7R | GGTCCTTTGGATCAATTTCC | 1336 | |
| A8R | TCATCTTCAGACATGGAGCC | 1454 | |
| 10R | GGTGAAGGAACTGCTTTGAC | 1690 | |
| A3R | GGAAATTTGGAGCCTACTGAG | 1811 | Bio[1] |
| 11R | TGTCAGATTTGCTGCTGAAC | 1870 | |
| E3R | CAAGAAGCTTATTGGCGAGC | 2132 | |
| ERB2 | ATTGAGTGGGAGCACAAAGG | 2224 | Bio[1] |

[1]Biotinylated at the 5' end.
[2]Enzyme site incorporated into the 5' end of the primer sequence.

## 2.18 Other PCR primers

**Table 2.3 Other PCR primers used in this study**

| Name | Locus | Sequence | Reference |
|---|---|---|---|
| 426F<br>427R | DXS178 | AACTGTGGATCCATTTCTGGCAGG<br>GAGACCAGTGTGCATCCAGTAAG | M.A.-J. O'Reilly personal communication |
| ALE1<br>ALE3 | ALU | GCCTCCACCAGTGCTGGGATTACAG<br>CCACTGCACTCCAGCCTGGG | Cole *et al.* 1991 |
| HYAC-C | YAC vector | GCTACTTGGAGCCACTATCGACTACGCGAT | P.de Jong personal communication |
| GLx6+<br>GLx6- | α GAL exon 6 | GGATGCTGTGGAAAGTGGTT<br>GGCCCAAGACAAAGTTGGTA | J. Davies personal communication |
| A1<br>A2 | λgt10 | AGCCTGGTTAAGTCCAAGCTG<br>CTTCCAGGGTAAAAAGCAAAAAG | Patanjali *et al.* 1991 |
| C1<br>C2 | λgt10 | CCACCTTTTGAGCAAGTTCAG<br>GAGGTGGCTTATGAGTATTTC | Parimoo *et al.* 1991 |

# CHAPTER 3
# CONSTRUCTION OF YEAST ARTIFICIAL CHROMOSOME CONTIG IN PROXIMAL Xq22

## 3.1 Introduction

At the start of this study, the XLA locus had not been mapped precisely but had been localised to the long arm of the X chromosome, at Xq22, by genetic linkage studies (Kwan *et al.* 1986; Mensink *et al.* 1986; Malcolm *et al.* 1987; Arveiler *et al.* 1987; Guioli *et al.* 1989; Kwan *et al.* 1990). A linkage study completed during the course of this study, in this laboratory (Lovering *et al.* 1993a), identified DXS442 and DXS101 to be new proximal and distal flanking loci, respectively, for the disease with the following consensus order of loci: centromere-DXS3-DXS366-DXS442-(DXS178-DXS265-XLA)-DXS101-DXS94-telomere (Barker *et al.* 1991; Lovering *et al.* 1993a). The critical region in which the XLA locus could lie was therefore refined to a genetic distance of 2-4cM (Barker *et al.* 1991). DXS178 was found to have no recombinations with XLA in seventy informative meioses and had a cumulative two point LOD score in excess of thirty, indicating very close linkage with the disease.

As part of a positional cloning approach to the isolation of the gene responsible for XLA in this laboratory, physical mapping techniques had been used to map probes specific for loci in Xq22. These studies revealed a number of CpG islands around the DXS178 locus, suggesting the presence of transcribed sequences (O'Reilly *et al.* 1992). The frequency of rare-cutting enzyme restriction sites in these islands, however, prevented the physical linkage of DXS178 with other probes from the region (O'Reilly *et al.* 1993a). In 1992, Allen and Belmont reported the presence of a CA repeat, DXS178 $(CA)_n$, which had been identified in a DXS178 positive YAC and was within 200kb of the DXS178 locus (Allen & Belmont, 1992). DXS265 was mapped to within 5kb of DXS178 by restriction mapping (Lovering *et al.* 1993b). Parolini *et al.* (1993) reported the order of these loci around DXS178 to be DXS178-DXS265-DXS178$(CA)_n$, although these loci

were not orientated relative to other loci in the region.

Using the approximation that 1cM genetic distance equals 1Mb physical distance (Hartley *et al.* 1984), the 2-4cM region between DXS442 and DXS101 equals around 2-4Mb of DNA. Yeast artificial chromosomes (YACs) allow the cloning of fragments of foreign DNA of up to 1-2Mb, and were therefore the most suitable vectors to use to clone DNA from a region of this size. The first aim of this study was, therefore, to construct a YAC contig across the XLA critical region, using probes specific for the two flanking loci, which had been defined by linkage analysis, and the loci around DXS178 to map the clones. This approach ran concurrently with a physical mapping study by Angela Sweatman, in this laboratory. Using the map of the contig, the second aim of this study was to select a YAC which was not chimeric and which covered the central part of the critical region for this disease, to use in a screen for candidate genes for XLA.

After the majority of the YAC analysis was complete and cDNA screening using one of the identified YACs was underway, the gene for XLA was identified (Vetrie *et al.* 1993c; Tsukada *et al.* 1993). This locus and two loci found to map very close by (*GLA* and 5D8) were incorporated into the map of the YAC contig. The results are included here for completion.

## 3.2 Results

### 3.2.1 Selection of YACs

YACs were selected on the basis of being positive for probes specific for loci from the XLA critical region (DXS442, DXS178 and DXS101), starting with DXS178, as this was the locus which mapped closest to the XLA locus by genetic linkage analysis. Eleven YACs, 178-A to 178-K, were a gift of Dr D.Vetrie, UDMS, London, and were all isolated on the basis being positive on an initial screen of the ICRF YAC library using a probe specific for DXS178. Primers 426F and 427R (section 2.18) were used to generate a DXS178 specific PCR product which was used as a probe on a membrane

containing these YACs. This probe was found to hybridise strongly to YACs 178-J and 178-K (Figure 3.1), allowing their positive identification as DXS178 containing YACs. These PCR primers and conditions were sent to the Human Genome Mapping Project Resource Centre in Harrow, Middlesex to screen the ICI YAC library, resulting in the identification of one positive YAC, 9DA6. The other YACs used in this study were provided by Dr D. Vetrie and Dr A. Monaco, as detailed in section 2.9.1.

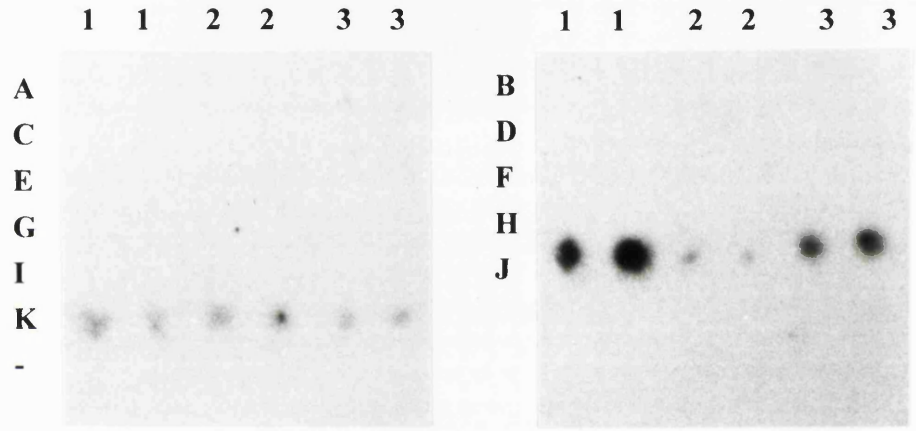
### 3.2.2 Sizing of YACs

All YACs were analysed by PFGE prior to Southern blot analysis. The sizes of the YACs (Table 3.1) were estimated by comparison with *S. cerevisae* marker or *H. wingei* chromosomes, after hybridisation of the membranes with probes specific for loci found in the clones. The DXS178 negative YAC was sized using a vector specific probe, as detailed in section 2.9.3. YACs 178-K, 442-A and DXS178 negative appeared to be unstable as analysis of DNA from these clones showed smaller bands which probably were produced by deleted derivatives of the YAC. An example of a pulsed field gel and its Southern blot analysis are shown in Figure 3.2.

### 3.2.3 Analysis of YACs 178-J and 178-K

YACs 178-J and 178-K were both found to be about 270kb in size (Figure 3.2b and Table 3.1). Only one YAC was present in each of the clones as determined by Southern blot analysis using probes specific for the YAC vector arms, but 178-K was found to have a faint deletion derivative of approximately 200kb. These two YACs were analysed by *Alu* PCR fingerprinting (Nelson *et al.* 1989) to characterise the cloned DNA. This technique involves the PCR amplification of inter-*Alu* fragments and is described in section 1.1.2. The results of this analysis are shown in Figure 3.3. YACs 178-J and 178-K were found to have identical product patterns suggesting that they were the same clone and, therefore, only YAC 178-J was used for further studies. The pattern seen in YACs 178-J and 178-K showed some products in common with YAC 9DA6 and no products in common with the DXS178 negative YAC. This result indicated that the insert contained in YAC178-J and YAC178-K may overlap with the insert contained in

**Figure 3.1 Secondary screen for DXS178 positive YACs**




Dot blot analysis of YAC DNA hybridised with a DXS178 specific probe. The letters down the side indicate the name of the clone. Numbers along the top indicate independent clones of the same YAC, in duplicate.

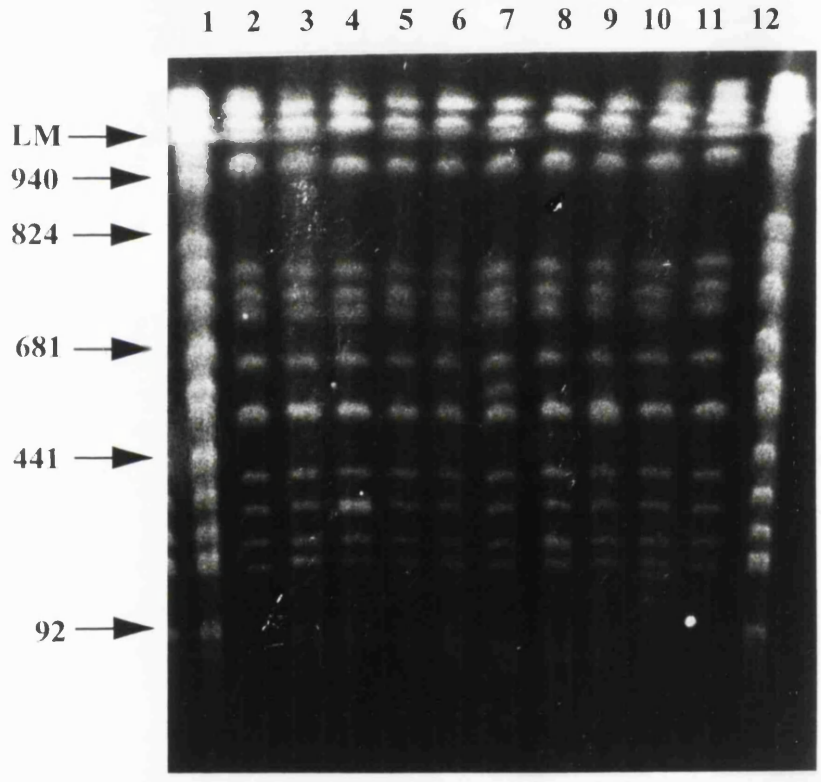
**Table 3.1 Sizes of YACs used in this study**

| YAC clone | Locus used for selection | Approximate size (kb) |
|---|---|---|
| 178-1 | DXS178 | 770 |
| 178-2 | DXS178 | 1200 |
| 178-3 | DXS178 | 570 |
| 178-4 | DXS178 | 250 |
| 178-5 | DXS178 | 550 |
| 9DA6 | DXS178 | 340 |
| 178-J | DXS178 | 270 |
| 178-K | DXS178 | 270<br>200 |
| 178-S | DXS178 | 230 |
| 178-M | DXS178 | 170 |
| 101-2 | DXS101 | 1300 |
| 101-8 | DXS101 | 900 |
| 101-5 | DXS101 | 470 |
| 442-A | DXS442 | 395<br>300 |
| 178 -ve | none | 200<br>150 |

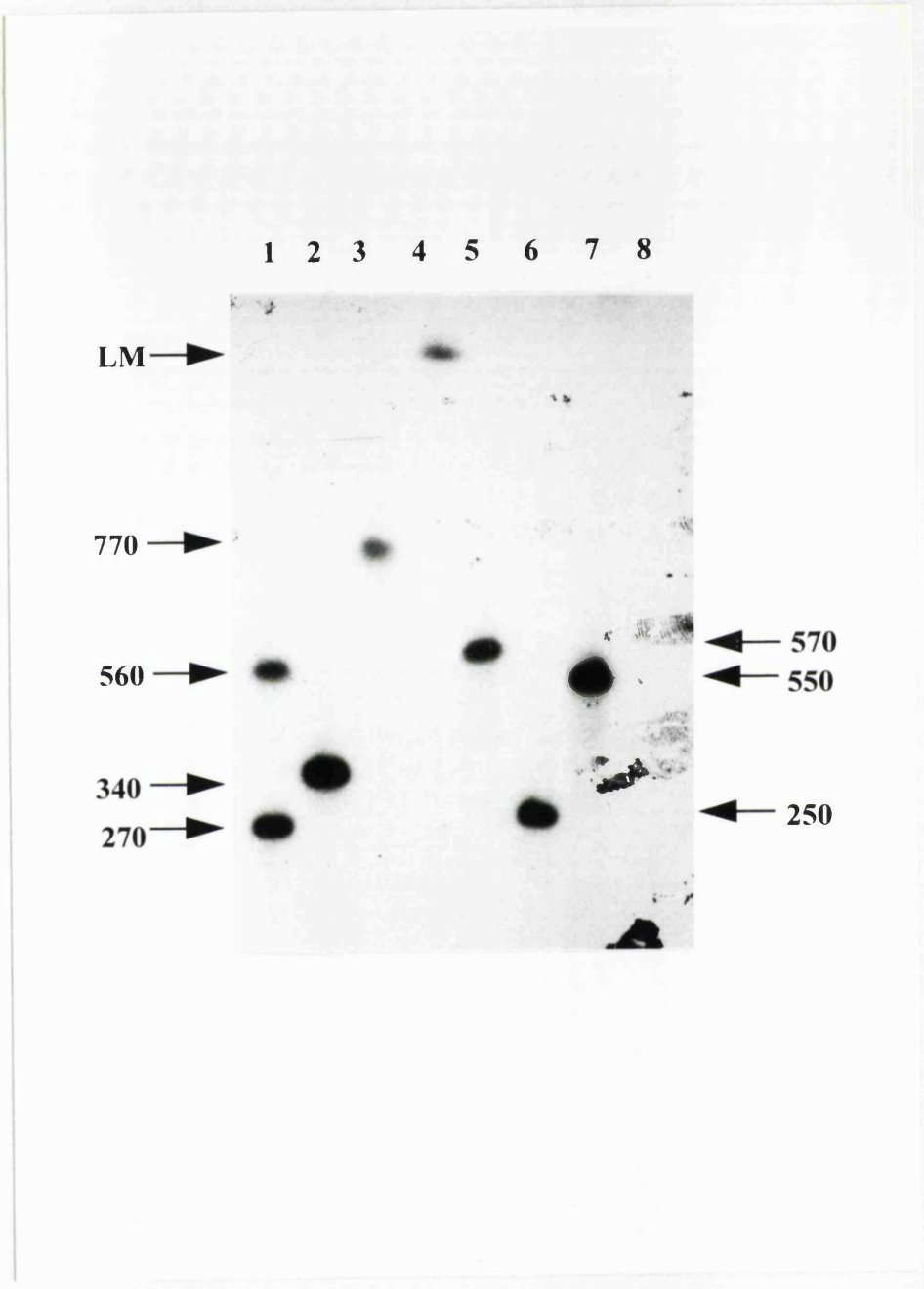
Approximate sizes of YACs used in this study, as assessed by PFGE, Southern blot analysis and comparison with *S. cerevisae* chromosomes. Two sizes indicate the presence of two positive YACs, the smaller one being most likely a deletion derivative.

**Figure 3.2 Analysis of YAC clones**

(a)



(b)




(a) PFGE of YAC DNA. Tracks are (1) YP148, (2) AB1380, (3) 178-J, (4) 9DA6, (5) 178-1, (6) 178-2, (7) 178-3, (8) 178-4, (9) 178-5, (10) DXS178 negative control YAC, (11) AB1380 and (12) YP148. *S. cerevisae* chromosome sizes are indicated by arrows, and (b) Southern blot analysis of part of the same gel of YAC DNA, hybridised with a DXS178 specific probe. Tracks are (1) 178-J, (2) 9DA6, (3) 178-1, (4) 178-2, (5) 178-3, (6) 178-4, (7) 178-5, (8) DXS178 negative. The 560kb band in YAC 178-J was present only in a second DNA preparation, indicating that recombination with a yeast chromosome may have occurred during propagation.The sizes of the DXS178 positive YACs are indicated by arrows. LM indicates limiting mobility.

YAC 9DA6, as would be expected as they were both selected using the same probe, and that neither of these YACs appeared to overlap with the DXS178 negative control YAC.

**Figure 3.3 Results of *Alu* PCR fingerprint analysis of YACs 178-K and 178-J.**


Gel shows (1) 1kb ladder size markers, (2) 178-K, (3) 178-J, (4) 9DA6, (5) DXS178 negative control YAC, (6) *S.cerevisae* AB1380 DNA, (7) no DNA control. Sizes of selected 1kb ladder bands are given in bp.

### 3.2.4 Sequential hybridisation of probes to YACs

#### 3.2.4.1 Analysis using DXS101

The DXS101 specific probe cX52.5 recognises five copies of a particular sequence, all copies of which map to Xq21.3-Xq22, and one of which contains a polymorphic *Msp*I site (Hofker *et al.* 1987). This polymorphism has been used in genetic linkage studies on XLA families and found to be the distal flanking marker for the XLA locus (Vetrie *et al.* 1993c). In order to make this probe more useful in genetic and physical mapping studies, single copy fragments were isolated from the cX52.5 plasmid and from two cX52.5 positive cosmids, as detailed in O'Reilly *et al.* (1993b). GPS is a 1.1kb *Sal*I/*Pst*I fragment isolated from the cX52.5 plasmid and was found to map between 460-540kb distal to DXS178 (O'Reilly *et al.* 1993b). Probe B1.4 is a 1.4kb *Bam*HI fragment isolated from a DXS101 positive cosmid, which was found to recognise the *Msp*I

polymorphism and to give a much cleaner signal than the whole cX52.5 plasmid, making it useful for genetic linkage analysis (O'Reilly *et al.* 1993b). A 700bp *Bam*HI fragment isolated from the same cosmid was subcloned and sequenced, enabling the design of PCR primers (O'Reilly *et al.* 1993b) which could be used to amplify a 550bp product (B550). This copy of DXS101 maps near the proteolipid protein (PLP) locus which is distal to DXS178 and XLA (O'Reilly *et al.* 1993a). A report published since the completion of this study (Vetrie *et al.* 1993b) has confirmed this data and mapped the B550 copy of DXS101 approximately 300kb distal to DXS24, and approximately 2Mb distal to DXS178. DXS101 specific probes GPS and B1.4 were used in hybridisations with the YACs in this study.

### 3.2.4.2 Results of hybridisations

YAC DNAs were initially studied by Southern blot analysis using six probes specific for the loci in the region between DXS366 and DXS101 (DXS366, DXS442, DXS178, DXS265, DXS101 (GPS) and DXS101 (B1.4), as detailed in Appendix I), and by PCR at the DXS178$(CA)_n$ locus (Allen & Belmont, 1992). An example of Southern blot analysis using a probe specific for DXS178 is shown in Figure 3.2a and a summary of the results are shown in Table 3.2. After this study had been completed, the gene for XLA was isolated (*BTK*) and loci mapping very close to the gene (5D8, *GLA*) were described (Vetrie *et al.* 1993c). These loci were included in this analysis to enable a more complete physical map of the YAC contig to be created, and the results are included here for completion.

The study of YACs 178-1 and 178-2 allowed the unambiguous orientation of three loci, DXS178, DXS265 and DXS178$(CA)_n$ which had been ordered but not mapped relative to other markers in Xq22 (Parolini *et al.* 1993). The presence of the DXS178 $(CA)_n$ locus in YAC 178-1 and its absence from YAC 178-2 enabled DXS178-DXS265-DXS178(CA)n to be orientated, with DXS178 being the most proximal (Table 3.2). The following order of loci could be assigned: cen-DXS442-DXS178-DXS265-DXS178$(CA)_n$-DXS101-tel, with the XLA locus lying between DXS442 and DXS101.

YAC 178-2 was positive for all the loci contained within the region from DXS366 to DXS265 (Table 3.2), so it could be concluded that the physical distance between these two loci is less than the length of the YAC, *i.e.* 1.2Mb. Since YACs can be unstable, however, and only one YAC spanned these loci, this data could not be considered definitive. A requested screen of the ICRF YAC library using a DXS442 specific probe identified only an unstable clone (442-A) of approximately 400kb. This was found to be only positive for DXS442 and could not therefore confirm the physical distance between any of the loci. All YACs were negative for the loci which flank this region - DXS94 (D.Vetrie, personal communication) and DXS3 (results not shown).

This analysis showed that the contig covered the region from DXS366 to DXS178, but it could not be shown that these YACs overlapped with any of the DXS101 positive YACs using these probes. Further probes, and possibly further YACs, would be required to complete the contig of the XLA critical region.

The subsequent inclusion of 5D8, *BTK* and *GLA* loci in the mapping of this contig showed that YAC 178-5 must contain a deletion, as it was found to be negative for *GLA* but positive for both the known flanking loci, *BTK* and DXS178. The definitive order of these three loci had been demonstrated by physical mapping (Vetrie *et al.* 1993a; Vetrie *et al.* 1993c).

### 3.2.5 Generation of YAC insert end probes

In order to link the YACs across the region between DXS178 and DXS101, it was necessary to generate end probes from the YAC inserts, either to show overlap or to initiate a screen for more clones. The method selected to do this was vectorette PCR (Riley *et al.* 1990). This technique requires the YAC DNA to be digested with a restriction enzyme and ligated to a vectorette, creating a vectorette library. The vectorette consists of a long pair of oligonucleotides (52bp plus overhang for ligation to a restriction enzyme cut "sticky end" if required) which are only homologous at each end (for 11 or 12 bp). When the oligonucleotides are annealed to one another, a characteristic bubble is created. All fragments of the YAC DNA are therefore present

**Table 3.2 Summary of the results of Southern blot analysis of YACs with Xq22 specific probes**

| YAC | DXS 366 | DXS 442 | * 5D8 | * *BTK* | * *GLA* | DXS 178 | DXS 265 | 178 $(CA)_n$ | DXS 101 GPS | DXS 101 B1.4 |
|---|---|---|---|---|---|---|---|---|---|---|
| 178-1 | - | - | + | + | + | + | + | + | - | - |
| 178-2 | + | + | + | + | + | + | + | - | - | - |
| 178-3 | - | - | + | + | + | + | + | + | - | - |
| 178-4 | - | - | - | - | - | + | + | - | - | - |
| 178-5 | - | - | + | + | - | + | + | + | - | - |
| 9DA6 | - | - | + | + | + | + | + | + | - | - |
| 178-J | - | - | - | - | + | + | + | + | - | - |
| 178-S | - | - | - | - | - | + | + | - | - | - |
| 178-M | - | - | - | - | - | + | - | - | - | - |
| 101-2 | - | - | - | - | - | - | nd | - | + | + |
| 101-8 | - | - | - | - | - | - | nd | - | + | + |
| 101-5 | - | - | - | - | - | - | nd | - | + | + |
| 442-A | - | + | - | - | - | - | nd | - | - | - |
| 178-ve | - | - | - | - | - | - | - | - | - | - |

+ indicates a positive signal, - indicates no signal and nd indicates not determined. Probes specific for the loci indicated were as listed in section 2.9.5. * indicates data produced after the *BTK* gene had been identified, but included here for completion.

in the library, with attached vectorette. In order to amplify specifically the insert DNA adjacent to the YAC vector sequence, *i.e.* the end of the YAC insert, the vectorette library is used in a PCR, with one primer homologous to the YAC vector sequence, and a second primer with sequence identical to the central part of the vectorette bubble. The template for the second primer can only be produced when the first strand synthesis from the YAC vector primer has been completed. Thus, only the DNA between the vectorette and the YAC vector sequence can be amplified, creating YAC insert end probes. After PCR amplification, the vector fragment could be removed by *Eco*RI digestion (left end 227bp, right end 174bp) to yield an insert specific probe.

End probes were generated from the YACs 178-1 and 178-2 using vectorette PCR on *Rsa*I and *Hae*III vectorette libraries. A summary of the results is shown in Table 3.3.

### 3.2.6 Identification of overlapping YACs using the end probes

The end probes were used in Southern blot analysis of all the YACs, to identify overlaps, and of somatic cell hybrid cell line DNA (AnLy (Zonana *et al.* 1988)). This cell line contains DNA from Xq12-qter as its only human DNA component, and therefore allows analysis of the genomic origin of the end probes. If the end probes are derived from this region of the X chromosome, it can be considered unlikely that the YACs were chimeric. The results of the hybridisations using the YAC end probes are summarised in Table 3.4. The results which indicated the YACs overlap are shown in Figure 3.4. The right hand end probe generated from the insert of YAC 178-1 (from the proximal end of the YAC insert DNA) hybridised to YAC 178-2, confirming the Southern blot analysis data which showed that the inserts in these YACs overlap. This right hand end probe did not hybridise to DNA from any of the other YACs.

The data shows that the probe generated from the left hand end of YAC 178-1 (at the distal end of the YAC insert) hybridised to YACs 101-2 on Southern blot analysis, indicating that these YACs overlap. This shows that by overlapping with YACs 178-2 and 101-2, YAC 178-1 completes the contig spanning the XLA critical region from DXS442 to DXS101. The data also shows that probes generated from both the right and

## Table 3.3 Summary of YAC end probes generated by vectorette PCR

| YAC | End | Library | Size of PCR product (bp) | Size of probe (bp) |
|---|---|---|---|---|
| 178-1 | right | *Rsa*I | 400 | 225 |
| 178-1 | left | *Rsa*I | 900 | 675 |
| 178-2 | right | *Rsa*I | 800 | 625 |
| 178-2 | left | *Hae*III | 350 | 125 |

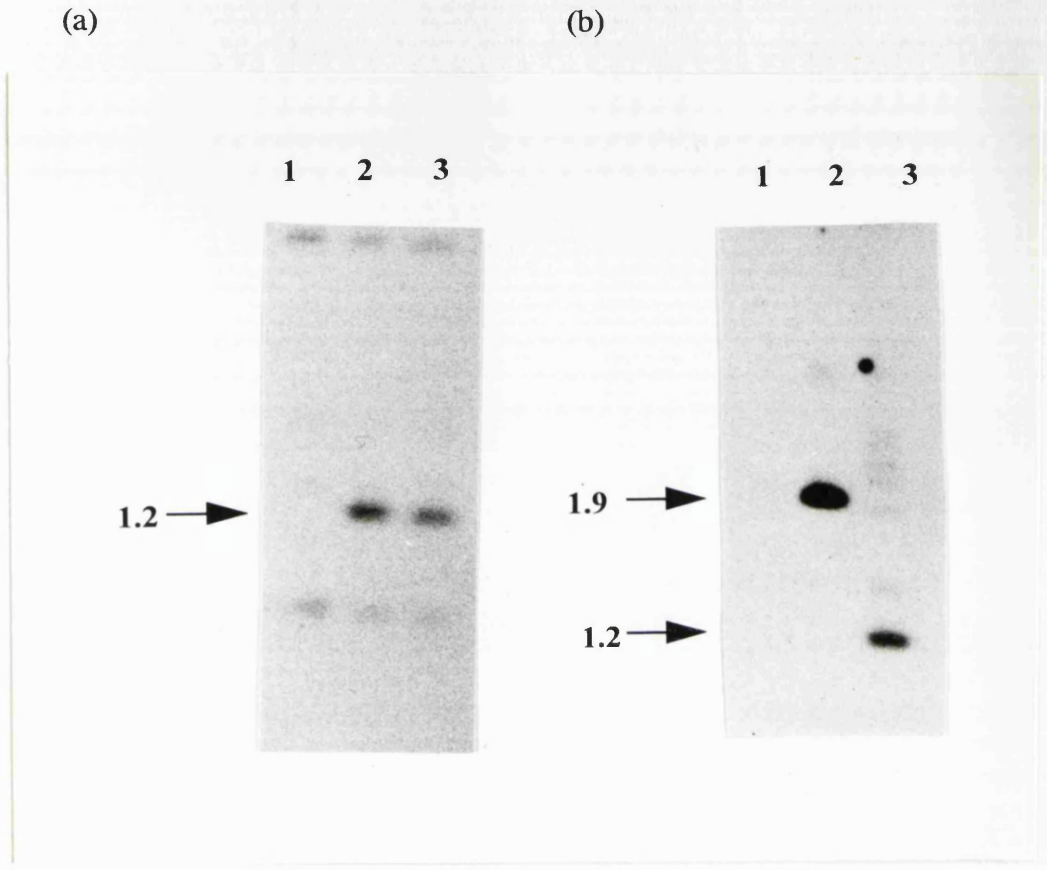
The YAC vector arm containing the ARS and centromeric sequences has been defined as the left arm and the arm without these sequences has been defined as the right arm. Sizes were estimated on an agarose gel using 1kb ladder DNA markers. Library names indicate the enzyme used in the creation of the vectorette library. The PCR product includes YAC vector sequences, which can be removed by digestion with *Eco*RI to give the probe.

## Table 3.4 Summary of analysis of Xq22 YACs using YAC 178-1 and 178-2 insert end probes

| YAC | End | YACs with positive hybridisation result | Hybridisation to AnLy |
|---|---|---|---|
| 178-1 | right | 178-1 and 178-2 (Figure 3.4) | repetitive |
| 178-1 | left | 178-1, 101-2,(Figure 3.4) 178-3, 178-5 | repetitive |
| 178-2 | right | 178-2 | negative? |
| 178-2 | left | 9DA6, 178-S, 178-J, 178-1, 178-2, 178-3, 178-4, 178-5 | positive |

End indicates vector orientation of YAC as detailed in the legend to Table 3.3. The presence of a band in the YAC DNA, which was absent in the negative control YAC DNA and the YAC host AB1380 DNA, was taken as a positive hybridisation signal. In some cases, YAC vector specific bands were found in all tracks containing YAC DNA, and these were discounted. This could be due to vector sequence contamination of the probe.

**Figure 3.4 Southern blot analysis using the YAC end probes**




Results of Southern blot analysis with (a) *Eco*RI digested DNA and using the YAC 178-1 right hand end probe, tracks contain (1) YAC 178-3, (2) YAC 178-2 and (3) YAC 178-1, and (b) *Msp*I digested DNA and using the YAC 178-1 left hand end probe, tracks contain (1)101-5, (2) 101-2 and (3) 178-1. Approximate fragment sizes are given in kb, measured against λ *Hind*III size markers.

left ends of YAC 178-1 hybridise to other YACs in the contig, suggesting that this YAC is not chimeric. These probes were fairly repetitive on human DNA and AnLy DNA, and the results of Southern blot analysis were unclear (results not shown).

The left hand end probe generated from the insert of YAC178-2 (from the distal end of the insert) was found to hybridise to many other YACs in the contig and the somatic cell hybrid line AnLy, but not to the parent cell line RAG, confirming that this end originates from Xq. The right hand end probe of YAC 178-2 (from the proximal end of the insert) only hybridises to itself, and not to any of the other YACs in the contig. The probe hybridises strongly and non-specifically to human genomic DNA and it was not possible

to determine if it specifically hybridised to the AnLy cell line DNA. It could not, therefore, be excluded that this YAC was chimeric. These probes were found to be small and contained repeat sequences and were therefore not useful in the physical mapping of cell line DNA (A.Sweatman, personal communication).

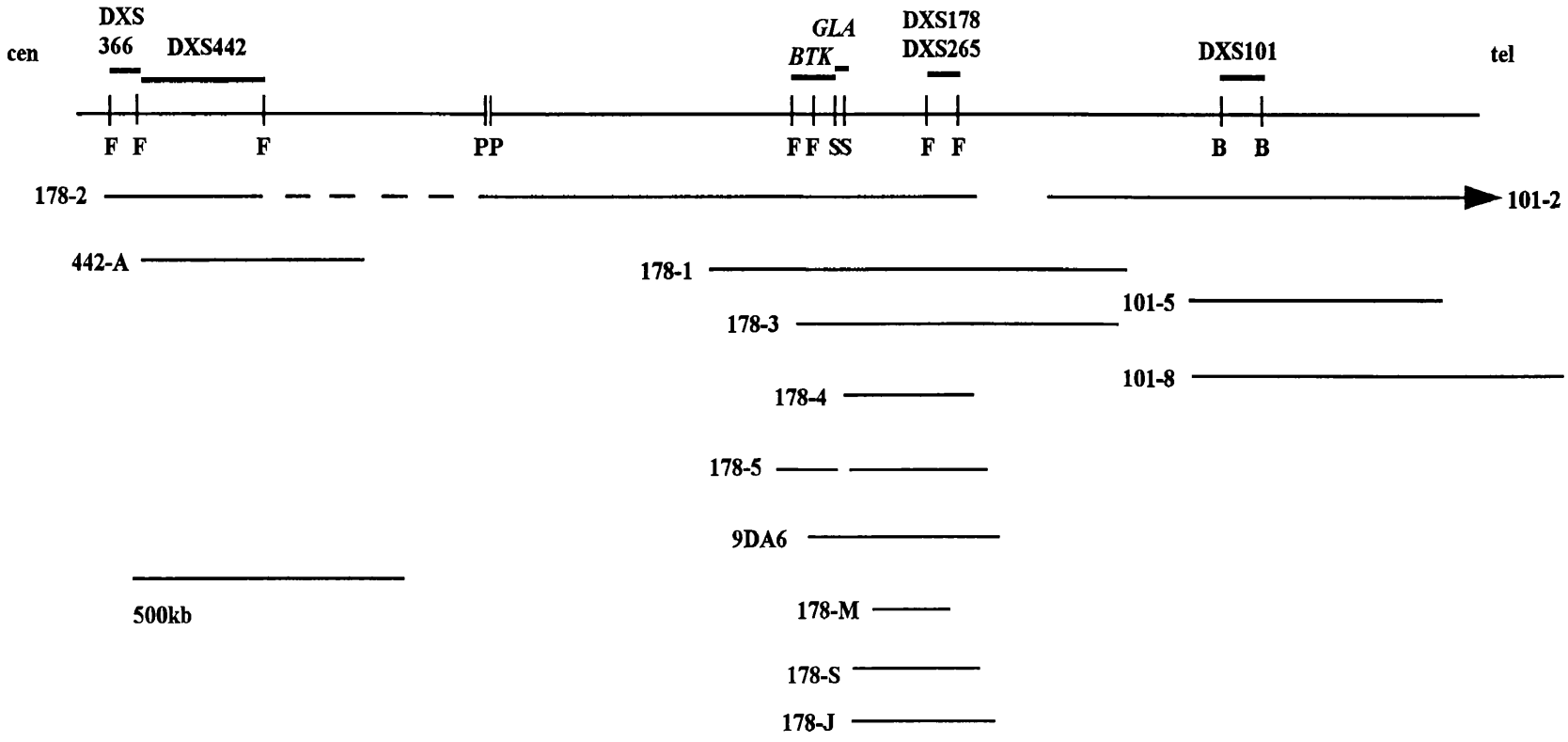
These results have shown, therefore, that YACs 178-2, 178-1 and 101-2 overlap and form a contig spanning at least 2.5Mb and covering the complete XLA critical region between DXS101 and DXS366 as shown in Figure 3.5. In addition to these three YACs, eleven smaller YACs were found to contain DNA from around DXS178, DXS442 and DXS101 and these were mapped relative to these loci and the other YACs.

### 3.2.7 Further analysis and orientation of probes mapping to YAC 178-2

PFGE and Southern blot analysis of cell line DNA in this laboratory did not show physical linkage between DXS442 and DXS178 (Sweatman *et al.* 1994), and therefore YAC 178-2 provided the only evidence that they could be physically linked. Other clones covering this region would be required to confirm this observation. As the only DXS442 positive clone isolated from the ICRF library did not extend distally to DXS178, it was decided to generate a restriction map of the YAC 178-2 to allow comparison with the physical map of the area. Using the information from YAC 178-2, DXS442 and DXS178 must be less than 1.2Mb apart.

In order to characterise YAC 178-2 further and to refine the distance separating DXS442 and DXS178 therefore, the YAC DNA was partially digested with the restriction enzymes *Bss*HII and *Sfi*I, and a restriction map produced. *Bss*HII and *Sfi*I were chosen for this analysis because they had both been used extensively in mapping studies of the area and *Sfi*I was known to cut between DXS366 and DXS442 (Sweatman *et al.* 1994). It is difficult to compare restriction maps of YAC DNA with maps of human cell line DNA because yeast DNA is not methylated, and therefore many of the methylation sensitive enzymes which are rare cutters in cell line DNA cut much more frequently in DNA which has been replicated in yeast. Restriction digestion with *Sfi*I is not dependent on methylation however and therefore the data may be more comparable with the cell line

## Figure 3.6 Representation of YAC contig across the XLA critical region from DXS442 to DXS101



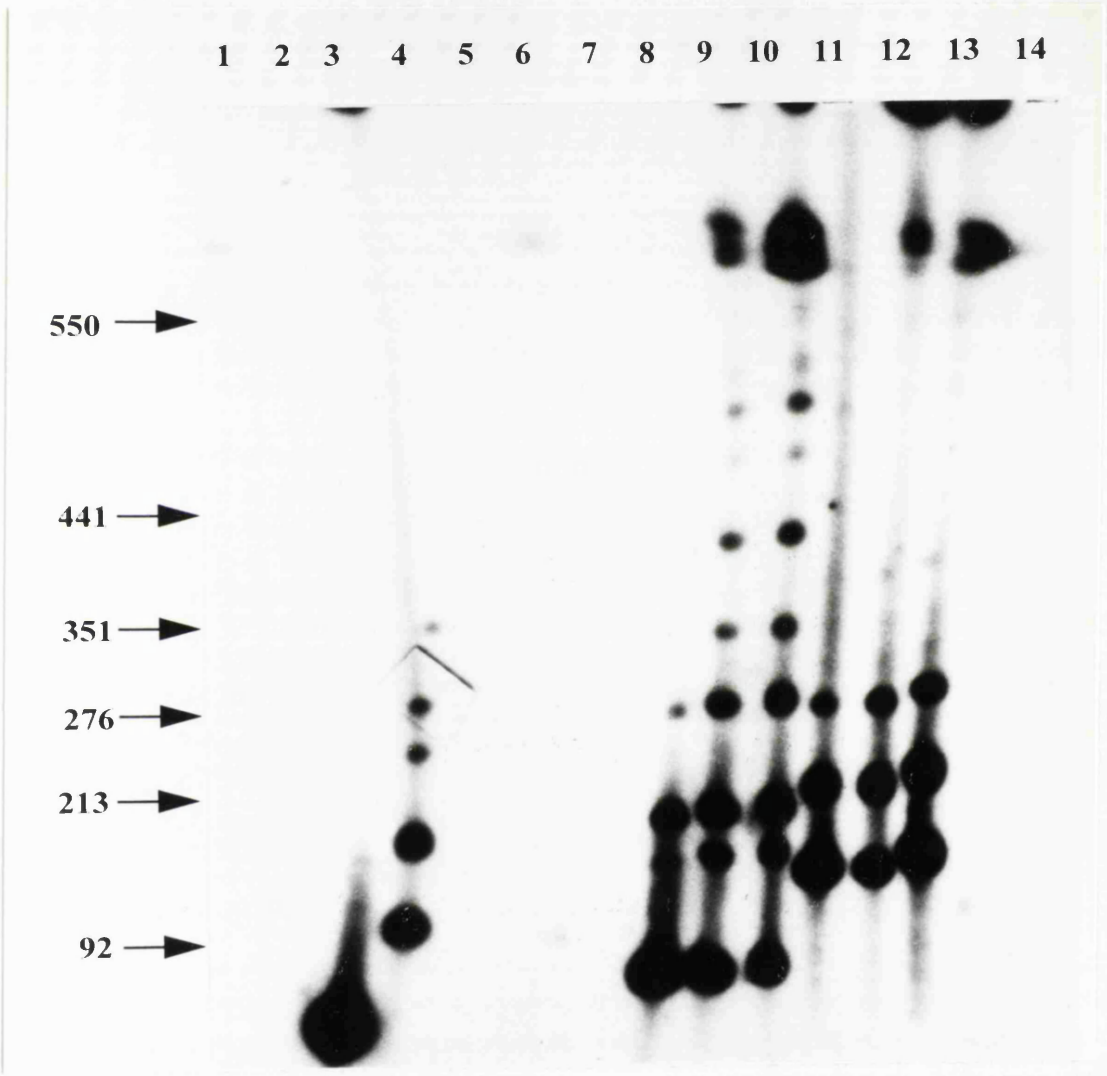

Schematic representation of YAC contig. YAC sizes are given in Table 3.2. Physical map as in A.K. Sweatman (1994). Only restriction fragments flanking the loci are shown. Gaps in YAC 178-5 and dashed lines in YAC 178-2 indicate regions of deletion. YAC 101-2 not drawn to scale. F = SfiI, P = PvuI, B = BssHII and S = SacII.

DNA mapping data. Restriction enzyme analysis of YAC 178-2 should also confirm the orientation of DXS366 and DXS442, with respect to the centromere and telomere, as their order relative to other Xq22 markers had been defined by a single recombination between DXS366 and XLA in an affected family (Barker *et al.* 1991), and they had not been physically orientated.

The complete and partial restriction fragments generated from YAC 178-2 were studied by Southern blot analysis using probes specific for loci known to be present in this YAC, the vector specific sequences and, after the *BTK* gene had been identified, for the genes *BTK* and *GLA*. The map was generated as the best fit of the data, incorporating what was known about the physical map of the region and maintaining consistency between the two sets of data. A 10% margin of error was allowed when adding up fragment sizes. An example of the Southern blot analysis is shown in Figure 3.6 and all the data is given in Appendix III. The restriction map is shown in Figure 3.7, with restriction fragments relevant to the discussion indicated.

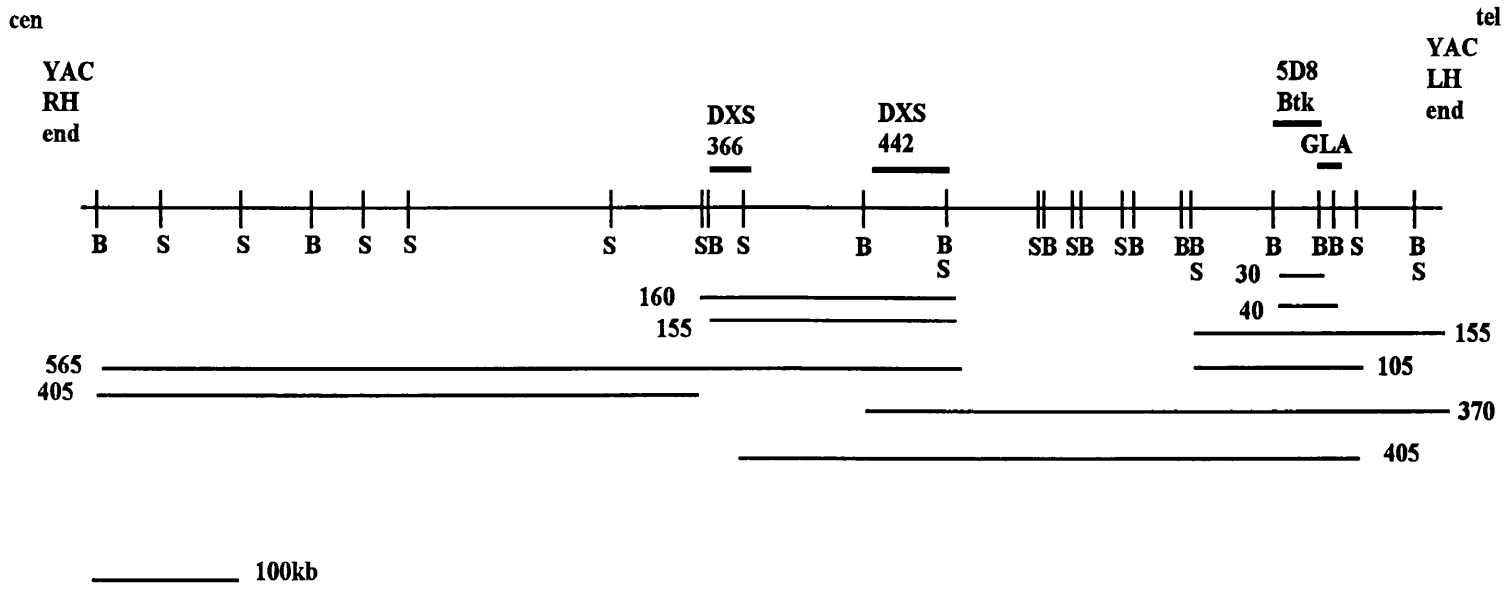
The data showed that DXS366 and DXS442 could be linked on a 160kb *Sfi*I fragment and a 155kb *Bss*HII fragment. These distances are compatible with, and indeed slightly reduce, the 280kb maximum distance between these two loci on the physical map produced in this lab (Sweatman *et al.* 1994) using human cell line DNA. The presence of a 405kb *Sfi*I fragment positive only with the right hand end probe suggested that DXS366 must be further than 405kb from the right hand end. DXS366 and DXS442 could be linked to the right hand end probe on a 565kb *Bss*HII fragment. DXS442 could be linked to the left hand end of the YAC on a 370kb *Bss*HII fragment and DXS442 and *BTK* were linked on a 405kb *Sfi*I fragment. *BTK* and 5D8 were linked on a 30kb *Bss*HII fragment and both of these loci were linked with *GLA* on 40kb *Bss*HII and 105kb *Sfi*I fragments. All three loci were linked to the left hand end of the YAC on 155kb *Bss*HII and 155kb *Sfi*I fragments. These fragment sizes are compatible with the physical map of this region produced in this laboratory (Sweatman *et al.* 1994), which suggested that *BTK* and the *GLA* loci were linked on a 100kb *Bss*HII fragment. 5D8 is now known to map within 7kb of *BTK* (Vorechovsky *et al.* 1994).

**Figure 3.6 Example of Southern blot analysis of YAC 178-2 partially digested with *Bss*HII and *Sfi*I**




Southern blot analysis of PFGE of YAC 178-2 DNA, hybridised with a DXS442 specific probe. Tracks contain (1) *S. cerevisae* YP148, (2) *Bss*HII complete digest (15U), (3) *Sfi*I complete digest (15U), (4) Blank, (5) YP148, (6) Blank, (7) *S. cerevisae* AB1380, (8) *Bss*HII (1U), (9) *Bss*HII (0.3U), (10) *Bss*HII (0.1U), (11) *Sfi*I (1U), (12) *Sfi*I (0.3U), (13) *Sfi*I (0.1U) and (14) YP148. Sizes of *S. cerevisae* YP148 chromosomes are given in kb.

## Figure 3.7 *Bss*HII and *Sfi*I restriction fragment map of YAC 178-2




Enzyme sites are shown as BssHII (B) and SfiI (S). Key partial digest fragments are shown. The YAC right hand (RH) end lies proximal and the YAC left hand (LH) end lies distal. All fragment sizes are given in Appendix III.

The restriction map of 178-2 between DXS442 and *BTK* generated by these partial digests is incompatible with the physical map of the region. The distance between *BTK* and DXS442 must be 1Mb or greater as there is a *Pvu*I site 400kb distal to DXS442 and another 600kb proximal to *BTK* (Sweatman *et al.* 1994). These two loci could not be linked by physical mapping even on *Pvu*I partial digests, whereas the maximum predicted physical distance between DXS442 and *BTK* from partial digestion of YAC 178-2 is 295kb. It must therefore be concluded that this YAC has a deletion of at least 705kb in between DXS442 and *BTK* and it cannot therefore be used to create a representative physical map of this region.

## 3.3 Discussion

### 3.3.1 A YAC contig spanning the XLA critical region

During this study, fourteen YACs were analysed using eleven probes specific for loci from the region spanning from DXS366 to DXS101 and were mapped into a contig which appeared to span the XLA critical region, from DXS442 to DXS101. Only one YAC (178-2) contained DNA from the region from DXS442 to DXS178 and further analysis indicated that this YAC did not faithfully represent the genomic DNA from that region. The flanking markers for the XLA critical region could not therefore be physically linked using the YACs from this study. Eight of these YACs (178-M,178-S, 101-5, 101-2, 178-K, 178-J and 442-1) have since been published as part of a 6.5Mb YAC contig in Xq22 (Vetrie *et al.* 1994), confirming the data presented here.

### 3.3.2 YAC 178-2 has an interstitial deletion

Further analysis of YAC 178-2, which contained DNA from DXS366 to DXS178, showed that the restriction map generated was not compatible with the known physical map of the region and it had to be concluded that there had been a deletion of at least 700kb from the central region of this YAC. The deleted region did not include any of the markers known to map to this region and was from the area between DXS442 and

5D8. The right hand end probe from YAC 178-1 would have been a useful marker to use in this analysis, but due to the instability of the vectorette library, it was not possible to generate more of this probe. The physical distance between DXS442 and DXS178 is not known, but must be 1Mb or greater (Sweatman *et al.* 1994). YACs have been found to have internal deletions in other studies (Foote *et al.* 1992; Palmieri *et al.* 1993), and it is thought that the deletion is produced by a transformation induced process. Care must be taken in all studies using DNA isolated in YACs to ensure that the clone is a true representative of the genomic DNA.

### 3.3.3 Physical mapping around the XLA region

A large number of CpG islands have now been identified in this region (O'Reilly *et al.* 1992; O'Reilly *et al.* 1993a; Vetrie *et al.* 1993a; Vetrie *et al.* 1993b; Sweatman *et al.* 1994), each of which may be associated with a gene. Due to the number of restriction sites mapped in the unmethylated YAC DNA, it was not possible to identify each particular island in YAC 178-2. The mapping of this contig has allowed the orientation of the DXS178-DXS265-DXS178$(CA)_n$ group of markers with respect to the centromere and telomere, with DXS178 being the most proximal marker. The proximal location of DXS366 with respect to DXS442 has also been confirmed. The restriction map of 178-2 has produced smaller minimum distances than had been determined by previous physical mapping studies between DXS366 and DXS442 and between 5D8, *BTK* and *GLA*. As this data is from a single YAC already known to have a deletion, confirmation of the distances from another source is needed before these smaller distances can be considered as definitive. Each of these results, however, contributes to a more accurate map of this part of Xq22. The generation of new probes from the ends of the YAC inserts provides the starting point for a screen for further YACs from this region, which would enable these data to be confirmed and increase the coverage of the contig between DXS442 and *BTK*. Further mapping studies of Xq22 have since been published (Vetrie *et al.* 1994; Vetrie *et al.* 1993b), and these are in agreement with the data presented here. DXS442 and DXS178 still have not been physically linked and the true extent of the region remains unknown. Several disease loci have been genetically mapped to this region, including X-linked megalocornea (Bleeker-Wagemakers *et al.* 1991) and FG syndrome

(Zhu *et al.* 1991), and a number of cDNA clones have been isolated from around DXS178 (Vorechovsky *et al.* 1994). Further analysis of this region of Xq22 will be required to map and characterise these loci.

### 3.3.4 Identification of a YAC clone to initiate a cDNA screening study

This mapping study showed that YAC 178-1 was 770kb in length and was not chimeric. This YAC mapped to the central portion of the XLA critical region and contained the DXS178 locus, which was known to be very tightly linked with the disease. This YAC was therefore chosen to initiate a cDNA screen for candidate genes for XLA using a cDNA selection approach.

# CHAPTER 4
# THE DIRECT cDNA ENRICHMENT APPROACH TO THE ISOLATION OF CANDIDATE GENES FOR XLA

## 4.1 Introduction

The identification of transcribed sequences from a defined genomic region is still the rate limiting step in any positional cloning approach to the identification of a particular gene. Selecting transcribed sequences from a cDNA library with a genomic clone from the region of interest appears to be the most successful method, although this relies on using a cDNA library made from a tissue where the gene is known to be expressed. The ease of identification of a gene is dependant on the abundance of the transcript in question in the cell. The isolation of rarer transcripts, which may be present at only $1x10^{-6}$ of total transcripts, is technically difficult, although normalised cDNA libraries have been developed in which relative cDNA levels are equalised (Patanjali *et al.* 1991; Takahashi & Ko, 1994).

At the time of this study, direct screening of gridded cDNA libraries, using YACs containing DNA from the region of interest as probes, had been used successfully to identify genes from defined chromosomal regions (Wallace *et al.* 1990; Elvin *et al.* 1990). Recently, greater sensitivity of selection has been obtained using cDNA enrichment techniques with a YAC immobilised either on nylon membrane (Parimoo *et al.* 1991; Lovett *et al.* 1991) or using magnetic bead technologies with the streptavidin/biotin capture system (Morgan *et al.* 1992; Korn *et al.* 1992). These enrichment techniques have been shown to increase the number of copies of the cDNAs encoded in the YACs found in the final selected pool from between 1000 fold, using one round of enrichment, and 7000 fold, using two rounds of enrichment, with membrane immobilisation (Lovett *et al.* 1991; Parimoo *et al.* 1991), to 100,000 fold enrichment using two rounds of enrichment using the biotinylation method (Morgan *et al.* 1992).

As these techniques rely on the specificity of hybridisation for their success, steps must be taken to prevent repeat sequences interfering with the selection procedure. Approximately 70% of the human genome is single copy, with the rest being comprised of repetitive sequences (reviwed in Lewin, 1993). Tandem repeats can be separated into satellite sequences (up to several megabases of repeated sequence making up the bulk of the heterochromatin), minisatellite sequences (smaller blocks of repeated sequence, found dispersed throughout the nuclear genome and making up telomeric sequences) and microsatellite sequences (monotonic runs, dinucleotide and trinucleotide repeats found throughout the genome). Other repetitive DNA sequences found in the human nuclear genome are the interspersed repeats. These can be divided into short interspersed nuclear elements (SINEs), of which the most conspicuous example is the *Alu* family, and the long interspersed nuclear elements (LINEs), of which the *Kpn* repeat family is an example. A significant number of expressed sequences in cDNA libraries are thought to contain repeat sequences, some of which are translated, found in unspliced messages or found in the 3' untranslated region (Crampton *et al.* 1981). Finally, yeast DNA contains tandem ribosomal sequences and this is thought to interfere with cDNA selection procedures as often some yeast genomic DNA is purified along with the YAC. This has been reported to result large proportions of selected cDNAs having ribosomal origin (Morgan *et al.* 1992; Fan *et al.* 1993). All of these types of repeat sequence can produce insurmountable background problems in cDNA selection studies, as the technique relies on the specificity of hybridisation for its success. Preannealing using human and yeast DNA can prevent some of these repeats contributing to the hybridisation reaction but some low level repeats will still be present. One study has reported the use of blocking of the genomic DNA before selection (Parimoo *et al.* 1991), while another reported the blocking of the cDNA library itself (Lovett *et al.* 1991).

The critical region containing the XLA gene had been defined as a 2-4cM region between DXS442 and DXS101 by genetic linkage analysis (Lovering *et al.* 1993a). Mapping of YACs (Chapter 3) indicated that the physical distance between these two markers was approximately 2.5Mb or less, but it was not known at this time that YAC 178-2 most probably had an internal deletion. YAC 178-1 was sized at 770kb and thought not to be chimeric. This YAC mapped to the central region of the XLA critical area and contained

the DXS178 locus, which was known to have no recombinations with the XLA locus in over 70 informative meioses. On this basis, YAC 178-1 was chosen to screen a cDNA library with the aim of isolating candidate genes for XLA. Since the molecular defect in XLA appears to arise at the pre-B cell stage of B cell development (section 1.4.4), a Nalm 6 pre-B cell cDNA library was used. The use of magnetic bead technologies for this type of technique had not been extensively documented prior to this study, but preliminary studies appeared to indicate that it was effective (Dr F. Cotter, ICH, London, personal communication). The blocking and washing conditions used were based on a previously reported membrane immobilisation method (Parimoo *et al.* 1991).

## 4.2 Results

### 4.2.1 Selection of cDNAs

YAC 178-1 was photobiotinylated and the Nalm 6 cDNA library was hybridised to the YAC as described in section 2.10. Three different sets of prehybridisation conditions were used (1, 2 and 3, see section 2.10.3), incorporating varying amount of yeast and human blocking DNA. Selected cDNAs from all three hybridisation conditions were amplified using two rounds of PCR with nested primers, A1 and A2, and C1 and C2 (section 2.18), using standard PCR conditions (Section 2.4.2) and an annealing temperature of 58°C. The amplification resulted in a smear of DNA between 500 and 1000bp as detected on an agarose gel (results not shown).

### 4.2.2 Subcloning of PCR amplified selected cDNAs

The PCR amplified selected cDNAs were subcloned into Bluescript KS plasmid as described in section 2.10.6, at a transformation efficiency of $6.92 \times 10^5$ transformants/$\mu$g. The X-Gal system (see section 2.10.9) allowed the selection of a total of 200 white recombinants, which were then plated out in duplicate onto gridded filters, with reactions 1, 2 and 3 being gridded out onto plated 4, 5 and 6, respectively, and grown on selective media, as described in section 2.10.9.

### 4.2.3 Primary screening of recombinants

Filters on which the 200 recombinants had grown were prepared for hybridisation (Buluwela *et al.* 1989), and hybridised with YAC178-1 DNA, in order to determine if the selected cDNAs hybridised to YAC insert DNA, and with radioactively labelled selected cDNA, to establish the copy number of the clone in the selected library. Out of the 95 recombinants which grew on the selective media and which were white, indicating an insert, 18 hybridised only with the YAC DNA, 30 only with the labelled cDNA and 47 hybridised with both. An example of the autoradiographs of the filters is shown in Figure 4.1. At this stage there did not appear to be any difference in the recombinant colonies originating from the cDNAs selected in the different hybridisation reactions. From these 95 colonies, 48 were initially selected for further analysis. These were picked so as to include a range of strengths and combinations of positive hybridisation; 13 hybridised to the YAC only, 12 hybridised to the selected cDNA only and 23 hybridised to both. The hybridisation results of the 48/95 recombinant colonies selected for a secondary screen are shown in Table 4.1. The results of the remainder are not shown.

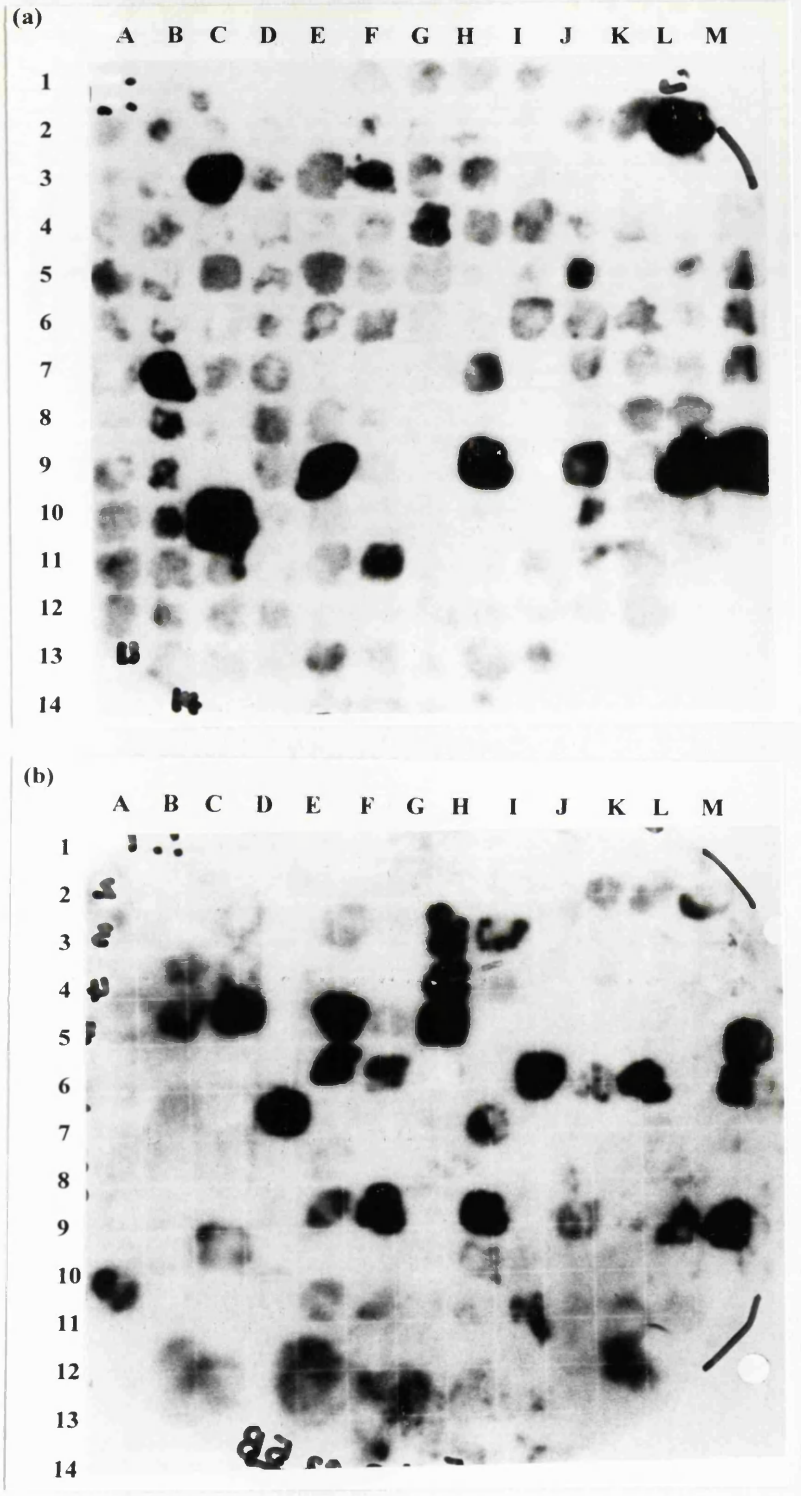
### 4.2.4 Secondary screening of recombinants

The inserts in these 48 recombinants were then analysed as described in section 2.10.12. Approximate sizes of inserts are given in Table 4.1. An example of a minigel showing digested recombinant plasmids is shown in Figure 4.2.

### 4.2.5 Analysis of inserts

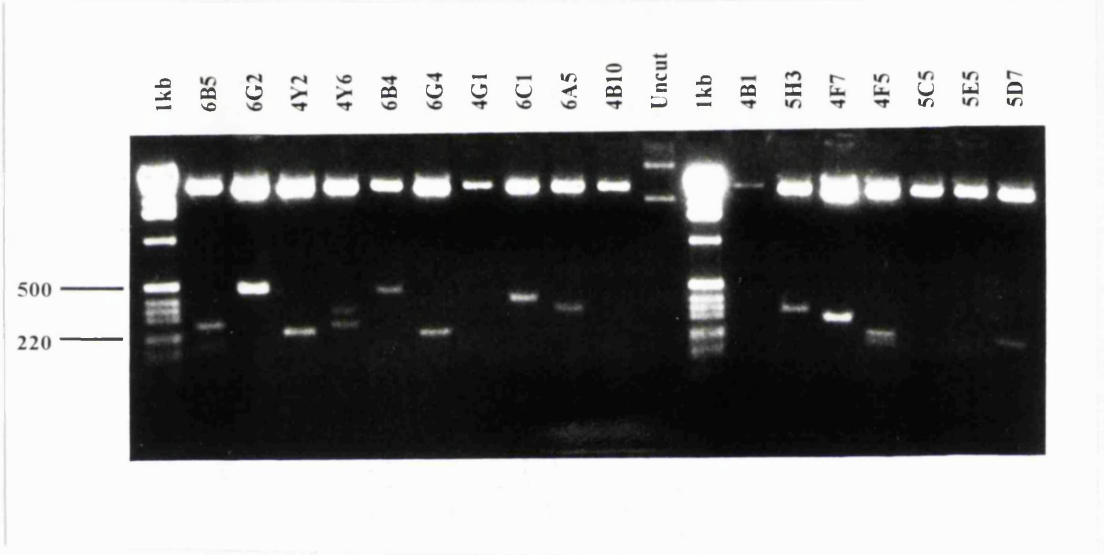
The inserts from the recombinant plasmids were systematically gel purified and hybridised to yeast (AB1380), YAC DNA and human DNA and to the other gel purified inserts, by Southern blot analysis. The results are given in Table 4.2 and examples of the autoradiographs are shown in Figure 4.3. The inserts from 16 recombinant plasmids were analysed before this work was terminated.

**Figure 4.1 Examples of hybridisations using the gridded colony filters**




Examples of autoradiographs of colony filter 5 hybridised with (a) radiolabelled YAC 178-1 DNA and (b) radiolabelled selected cDNA. Colonies were named according to filter number (4, 5 or 6) and grid reference (letter and number).

**Figure 4.2 Example of minigel of *Eco*RI digested recombinant plasmid preparations showing subcloned inserts**




Names of recombinants as described in legend to Figure 4.1. Sizes of selected 1kb ladder bands are given in bp. U indicates uncut plasmid. 1kb indicates marker track.

**Table 4.1 Summary of primary screening hybridisation results of the 48 colonies selected for the secondary screen, and their approximate plasmid insert size**

| Colony ref. | Hybridisation with YAC | Hybridisation with cDNA | Approx. size of insert | Colony ref. | Hybridisation with YAC | Hybridisation with cDNA | Approx. size of insert |
|---|---|---|---|---|---|---|---|
| 4X7 | + | ++ | 300 | 5G4 | + | ++ | 300 |
| 4X10 | ++ | ++ | 350 | 5G5 | - | ++ | 400 |
| 4Y2 | ++ | - | 250 | 5H3 | + | + | 350 |
| 4Y3 | + | - | 250<br>400 | 5H9 | ++ | + | 200 |
| 4Y6 | ++ | + | 270<br>350 | 5I4 | + | - | 450 |
| 4A3 | - | ++ | 270 | 5J9 | + | + | 600 |
| 4B6 | ++ | + | 450 | 5K6 | + | ++ | 300 |
| 4B7 | + | - | 450 | 5L2 | ++ | ++ | 200 |
| 4C1 | - | ++ | 200 | 5L9 | ++ | ++ | 600 |
| 4C2 | - | ++ | 370 | 5M7 | + | ++ | 400 |
| 4D3 | - | ++ | 240 | 6Y4 | + | + | 200 |
| 4F5 | + | - | 270<br>220 | 6Y5 | ++ | + | 350 |
| 4F7 | + | - | 300 | 6Y7 | + | ++ | 350 |
| 4I5 | - | + | 240 | 6A5 | - | ++ | 400 |
| 5B7 | ++ | - | 350 | 6B1 | + | - | 600 |
| 5C3 | ++ | - | 350 | 6B3 | + | + | 200 |
| 5C10 | ++ | + | 300 | 6B4 | ++ | ++ | 490 |
| 5D7 | - | ++ | 270 | 6B5 | ++ | ++ | 270<br>170 |
| 5E3 | + | + | 350 | 6C1 | + | ++ | 450 |
| 5E5 | + | ++ | 400 | 6E4 | + | - | 400 |
| 5E6 | + | ++ | 300 | 6F3 | + | ++ | 190 |
| 5E9 | ++ | + | 300 | 6G2 | ++ | - | 500 |
| 5F3 | + | - | 300 | 6G4 | + | ++ | 250 |
| 5F11 | ++ | - | 300 | 6G6 | ++ | ++ | 450 |

+ indicates positive signal, ++ indicates very strong signal and - indicates no signal. Approximate insert sizes are given in bp. Colony reference is as described in the legend to Figure 4.1

These results show that insert 6E4 was the only cDNA analysed which appeared to hybridise to relatively unique human DNA sequences (eight restriction fragments could be seen) and to the YAC. Inserts $4Y3_{250}$, $4Y3_{400}$, 5B7, 5J9, 5K6, 5L9 and 6B1 contain highly repetitive sequences, as they all hybridise extensively and non-specifically to total human DNA. These repetitive sequences may have resulted in these cDNAs hybridising to the YAC. Four other inserts (4X7, 5M7, 6G6 and 4F7) hybridised to a low number of restriction fragments in human DNA but did not appear to hybridise specifically to the YAC, whereas two other inserts (4D3 and 5G5) did not hybridise to either human DNA or the YAC.

Insert $4Y3_{250}$ is an example of an insert which was found to be highly repetitive (Figure 4.3a). This insert hybridised to multiple other inserts, and strongly and non-specifically to both the YAC DNA and human DNA. Many of the inserts which hybridised to $4Y3_{250}$ also hybridised to $4Y3_{400}$, suggesting that these inserts were part of the same repetitive sequence and were not the result of a coligation event in the preparation of the cDNA library.

Insert 4F7 is an example of a cDNA found to hybridise specifically to human DNA (one restriction fragment), but which did not hybridise to the YAC (Figure 4.3b), suggesting that this insert may be derived from a gene which is not located in the genomic region contained in the YAC. This clone was not, therefore, specifically selected in this procedure. The other inserts which were positive with 4F7 are not ones which appear frequently in the list of other subcloned cDNAs hybridising to labelled inserts, suggesting that this insert may contain a low copy number repeat.

Insert 6E4 was the only cDNA which hybridised to a limited number of restriction fragments in human DNA (although eight fragments were seen) and to the YAC DNA (Figure 4.3c), suggesting that this cDNA may have been selected because the gene from which it is derived, or a homologous gene, is contained in the YAC. This insert did not hybridise to any of the other 48 secondary screen inserts. In the primary screen, this insert hybridised to the YAC DNA but not to the labelled cDNA which confirms the suggestion that it was transcribed at a low level in this cell line.

Insert 5M7 hybridised to a single band in human DNA, but not to the YAC (results not shown). During the analysis of this cDNA, the labelled insert was hybridised to a Southern blot containing 4XY, XX and XY human DNA and to the somatic cell hybrid AnLy which contains part of the human X chromosome (Xq12 to Xqter) as its only human DNA on a murine background (Figure 4.4a). This insert hybridised to a single fragment in the human DNA and to a larger fragment in the AnLy DNA. When the blot was washed at increased stringency, this larger fragment in the AnLy DNA could no longer be seen (Figure 4.4b). This suggested that the 5M7 cDNA was hybridising to the murine homologue of the gene encoding this cDNA.

### 4.2.6 The identification of the gene responsible for XLA

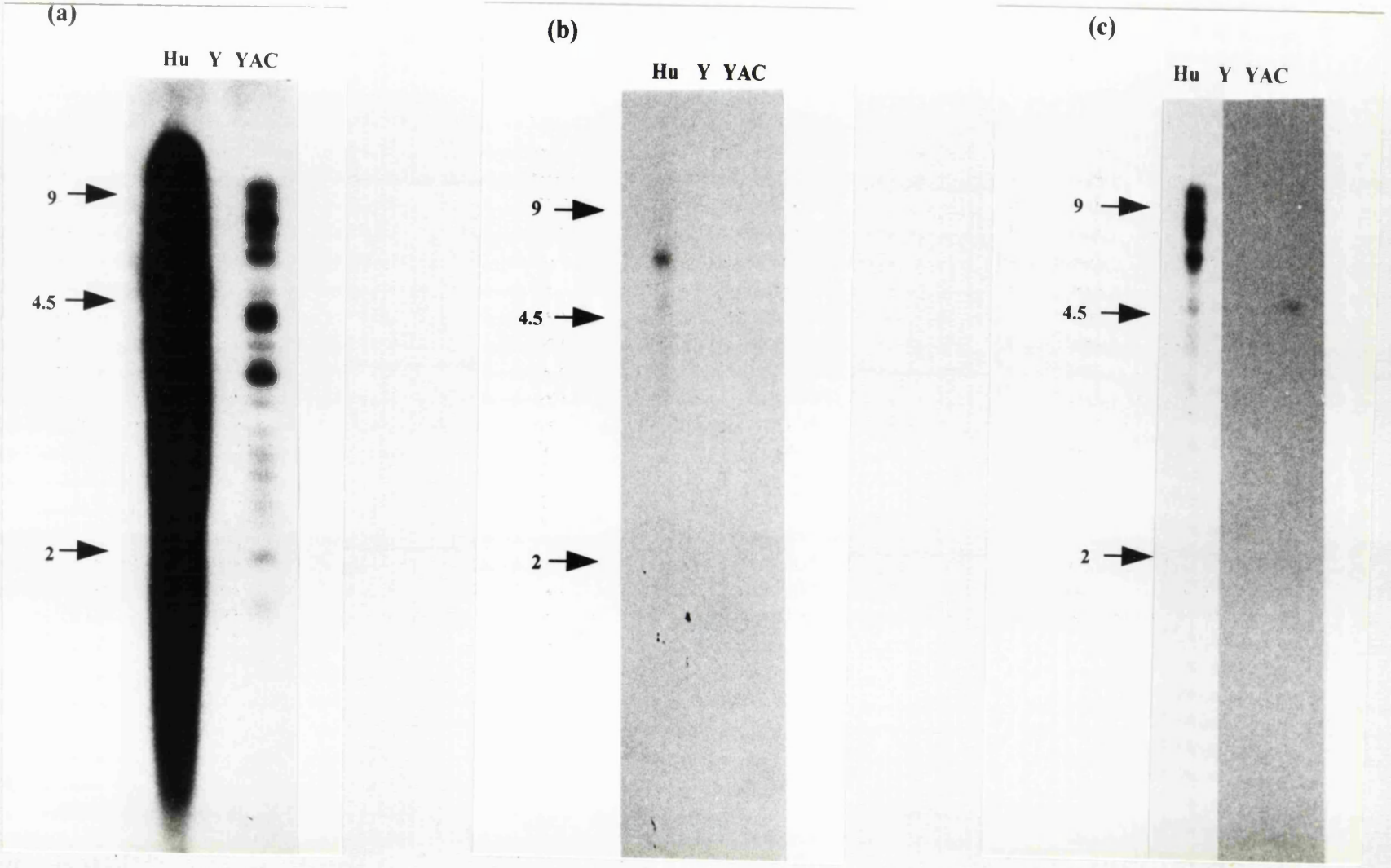
During this study, a strong candidate gene for XLA, *BTK*, was found by Dr David Vetrie (Guy's Hospital, London) (Vetrie *et al.* 1993c), using cDNA selection techniques. Vetrie *et al.* used direct selection of a cDNA library, made from an EBV positive Burkitt's lymphoma cell line, using a 640kb YAC which mapped to the XLA critical region immobilised on a nylon membrane. Mapping studies had localised the $\alpha$ galactosidase locus (*GLA*) to 70-140kb proximal of DXS178, the locus most closely linked with XLA (Vetrie *et al.* 1993a). As the 640kb YAC was positive for both the DXS178 and *GLA* loci, the *GLA* gene could be used as a positive control. This YAC was gel purified and immobilised on a nylon membrane before being hybridised with the cDNA library. The presence of the *GLA* gene on the YAC enabled the success of the cDNA enrichment procedure to be followed throughout the manipulations. The *GLA* cDNA was found to be enriched three-hundred fold after a single round of selection. The selected cDNAs were subcloned and automated griding facilities enabled 1536 colonies with inserts to be analysed. 104 of these subcloned cDNAs hybridised to thirteen cosmids from across the region and could be assigned to twelve groups. Southern blot analysis using one these groups of cDNAs showed genomic alterations in eight out of thirty-three XLA patients, including five showing genomic deletions of part or all of the gene, two of which appeared to be intragenic deletions. Sequencing identified two patients with point mutations. No alterations were found on Southern blot analysis of 150 normal X chromosomes. These results indicated a link between this group of cDNAs and the gene

## Table 4.2 Summary of Southern blot analysis using plasmid inserts as probes

| Colony ref. | YAC | Yeast | Human | Other subcloned cDNAs |
|---|---|---|---|---|
| 4A3 | - | - | +/-smear | 0.5 x SSC: 4C1, 4E2, 4E10, 4G1, 4X7, 5C3, 5C5, 5E5, 5E6, 5G4, 5H7, 5I6, 5K2, 6B5, 6F3, 6G2, 6G6, 6H2, 6H3, 6J3, 6J4, 6Y4 |
| 4D3 | - | - | - | 0.5 x SSC: 4B1, 4D3, 4F5, 4I5, 4X10, 5D7, 5E9, 6B3, 6G4 |
| 4F7 | - | - | 1 band | 0.5 x SSC: 4A3, 4D3, 4E1, 4I5, 4Y3, 5D7, 5E3, 5E9, 6B3, 6D4, 6G4 |
| $4Y3_{250}$ | ++ | - | ++ smear | 1 x SSC: 4B6, 4B7, 4C2, 4E8, 4Y2, 4Y6, 4X10, 5B7, 5C10, 5G4, 5H3, 5M7, 6A5, 6B4, 6B5, 6C1, 6G6 |
| $4Y3_{400}$ | ++ | - | ++ smear | 1 x SSC: 4B7, 4X10, 4Y2, 4Y6, 5C3, 5C10, 5M9, 6B4 |
| 4X7 | - | - | 1 band | 0.5 x SSC: 4A3, 4F7 |
| 5B7 | ++ | | ++ smear | 0.1 x SSC: 4B6, 4X10, 4Y6, 5C10, 5L2, 5L9, 5M9, 6B4, 6G6 |
| 5G5 | - | - | - | none |
| 5I4 | - | - | +/- smear | 1 x SSC: 4Y2, 4Y6, 5K6 |
| 5J9 | + | - | + smear | 0.1 x SSC: 4C2, 6A3, 6B5, 6C1, 6Y7<br>0.5 x SSC: 4B6, 4B7, 4C2, 4X10, 5E5, 5H3, 5M7, 6A3, 6B1, 6B5, 6C1, 6Y7 |
| 5K6 | - | - | + smear | none |
| 5L9 | + | - | + smear | 0.5 x SSC: 4X10, 4B6, 4Y2, 4Y6, 4B7, vector sequences |
| 5M7 | - | - | 1 band | 1 x SSC: 4B7, 4Y6, 5E6<br>AnLy +ve 1xSSC, -ve 0.5XSSC |
| 6B1 | + | - | + smear | 0.5 x SSC: 5H3, 5J9, 6A5, 6B5, 6C1, |
| 6E4 | + | - | 8 bands | none |
| 6G6 | - | - | 4 bands | 0.5 x SSC: 4A3, 4C1, 4E2, 4G1, 5B7, 5C5, 5D7, 5E5, 5G4, 5H7, 5I6, 5J9, 5K2, 5L2, 6B1, 6B5, 6D4, 6E4, 6F3, 6F4, 6H2, 6H3, 6J4, 6Y4 |

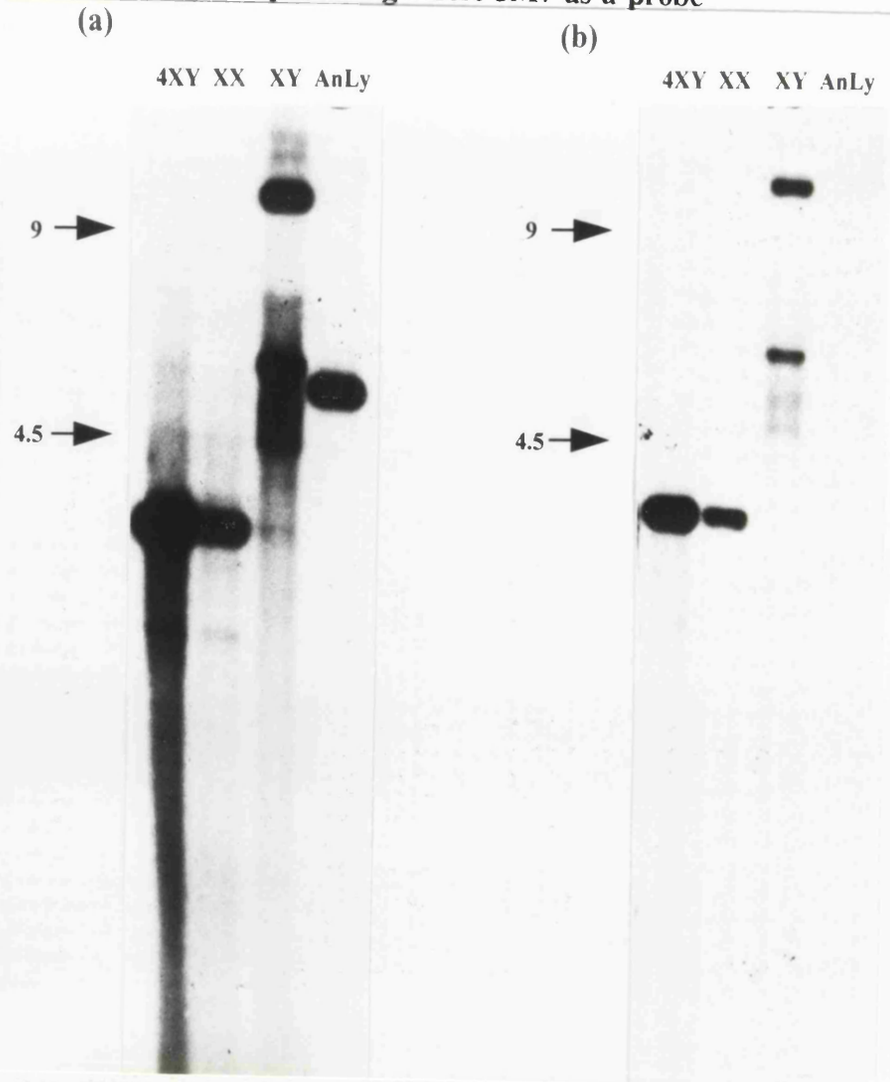
Results of hybridisation experiments using purified inserts as probes in Southern blot analysis of *Eco*RI digested DNA from YAC 178-1, total yeast (AB1380), human genomic and the other subcloned inserts. - indicates no hybridisation signal, +/- indicates weak hybridisation, + indicates a positive signal and ++ indicates a very strong signal. All of the 48 secondary screen inserts were screened with each radio-labelled cDNA. AnLy (see insert 5M7) is a somatic cell hybrid containing human Xq12-qter (Zonana *et al.* 1988). The results of the human DNA analysis are a visual description to indicate the degree of specificity of hybridisation. Colony references are as described in the legend to Figure 4.1. Where a subscript number is given after the colony reference, this is to indicate which of the fragments found in the insert is being used. Washing conditions are included to indicate the strength of the hybridisation.

**Figure 4.3 Autoradiographs showing Southern blot analysis using inserts $4Y3_{250}$, 4F7 and 6E4 as probes**




Southern blot analysis of YAC 178-1, *S. cerevisae* AB1380 and human genomic DNA probed with (a) 4Y3250, (b) 4F7 and (c) 6E4. Sizes of λ*Hind*III markers are given in kb.

**Figure 4.4 Southern blot analysis using insert 5M7 as a probe**




Southern blot analysis of 4XY, XX and XY human genomic DNA and AnLy somatic cell hybrid DNA using insert 5M7 as a probe washed at (a) 1 x SSC and (b) 0.5 x SSC. Sizes of λ*Hind*III markers are given in kb.

responsible for XLA.

The same gene was simultaneously isolated by Tsukada *et al.* (1993), who were aiming to identify genes involved in B cell growth and development. By screening a cDNA library produced from B cell progenitors with the kinase domain of the *ltk* gene at low stringency, they isolated a mouse cDNA. They then isolated the human homologue from a human library and showed that this gene mapped to the X chromosome by *in situ* hybridisation. Its position on the X chromosome and its expression in B cell progenitors made it a good candidate for the causative gene in XLA. Using Southern blot analysis, however, they were unable to detect mutations within this gene in the DNA of XLA patients. Using an antisera produced against part of the protein encoded by this gene, they did show, however, that there is little or none of this protein in some XLA patient B cell lines, supporting the argument that this gene is involved in XLA.

As there was strong evidence that this gene was in fact the causative gene in XLA, it was no longer appropriate to continue to screen the recombinant plasmid inserts selected in this study.

### 4.2.7 Success of the cDNA enrichment procedure in this study

After the publication of the work of Vetrie *et al.* (1993c), the YACs were analysed for the presence of the *BTK* gene, as shown in section 3.2.4. The gene was shown to be present in YAC 178-1, which was used for this screening study, and in four other YACs in the contig presented in Chapter 3, by Southern blot analysis (see Figure 3.5), using the *BTK* cDNA as a probe (provided by Dr D.Vetrie, London). The cDNAs isolated in this study were screened for the presence of the *BTK* cDNA by Southern blot analysis. The *BTK* gene was not detected in the 48 cDNA clones obtained in the secondary screen of this study. Technical difficulties prevented the analysis of the original cDNA library and the enriched pool of cDNAs for the presence of the *BTK* or the *GLA* cDNAs.

## 4.3 Discussion

### 4.3.1 The success of direct cDNA enrichment

Five out of the sixteen clones characterised in this study appeared to encode non-repetitive sequences (4F7, 4X7, 5M7, 6E4 and 6G6). Only one of these, however, appeared to map to the YAC (6E4). Seven out of the remaining clones appeared to contain repeat sequences as they hybridised strongly and non-specifically to a Southern blot of human DNA, and the final two did not appear to hybridise to YAC, yeast or human DNA. This indicated a specificity of selection of 1/16 (approximately 6%). Preannealing of non-specific sequences which may interfere with specific hybridisation of cDNAs to the YAC appears to be a critical part of direct cDNA selection. In this study, reactions 1 and 2 were pre-hybridised with yeast DNA and reaction 3 was prehybridised with human DNA and these reactions resulted in the subcloned cDNAs gridded on plated 4, 5 and 6, respectively. The subclones which originated from reaction 3 were slightly less repetitive than the clones originating from reactions 1 and 2 (see Table 4.2), suggesting that prehybridisation with human DNA plays an important role in the specificity of the hybridisation reaction. None of the reactions resulted in subclones which hybridised to yeast DNA, suggesting that interference from yeast ribosomal sequences was not significant in this experiment. Yeast sequences have been found to be a frequent contaminant in other studies (Morgan *et al.* 1992; Fan *et al.* 1993). The hybridisation of the inserts to the other sub-cloned cDNAs did not appear to show any particular pattern, but certain sub-clones were detected more frequently than others, for example 4X10, 4B7, 5C10 and 4Y6. These inserts tended to be detected by the hybridisation of labelled inserts which were found to be repetitive on human DNA, suggesting that these inserts contained repetitive sequences.

The results indicated that the cDNA selection procedures employed in this study gave a low level of enrichment of cDNAs mapping to YAC 178-1. Subsequent studies, after the *BTK* gene had been isolated, showed that the *BTK* gene was present on YAC 178-1 but failed to show the presence of *BTK* in the cDNAs selected in this study. Further blocking steps to prevent non-specific hybridisation and a second round of selection may

have been necessary to ensure the success of this technique. Attempts to isolate a $V_{pre\text{-}B}$ positive YAC to use as a positive control for this experiment, as the $V_{pre\text{-}B}$ cDNA was known to be contained in the cDNA library, were unsuccessful.

### 4.3.2 Comparison of methodologies for the isolation of the XLA gene

The aim of this study was to isolate the gene responsible for XLA using a direct cDNA enrichment procedure with a YAC 178-1, from a YAC contig containing the XLA critical region. Sixteen subcloned cDNA from the enriched pool were characterised before the publication of the *BTK* gene by Vetrie *et al.* (1993) and Tsukada *et al.* (1993). Vetrie *et al.* used a similar technique to the one described in this study. They used a 640kb YAC, similar in size to the 770kb YAC used in this study, and immobilised the YAC DNA to a nylon membrane. This method of immobilisation was more widely used at the time of this study (Parimoo *et al.* 1991; Lovett *et al.* 1991), but magnetic bead technologies, as used in this study, has now been shown to give equivalent or better results (Morgan *et al.* 1992; Korn *et al.* 1992). Knowledge of the presence of the *GLA* gene on the YAC allowed Vetrie *et al.* to use the *GLA* cDNA as a positive control. Preliminary *GLA* mapping data, from the study of somatic cell hybrids in this laboratory, had suggest that this locus mapped outside the XLA region (M.-A.J. O'Reilly, personal communication). Pre-hybridisation conditions have been shown to be critical for the success of this type of selection technique; Vetrie *et al.* prehybridised the YAC DNA and the cDNA with sheared placental DNA (Parimoo *et al.* 1991; Lovett *et al.* 1991), whereas in this study, only the YAC DNA was blocked with human and yeast DNA (Parimoo *et al.* 1991). Vetrie *et al.* used a cDNA library derived from an EBV positive Burkitt's lymphoma cell line and the study described in this thesis used a cDNA library derived from a Nalm-6 pre-B cell line, both of which have now been shown to express *BTK* at equivalent levels (Genevier *et al.* 1994a), suggesting that both libraries should have contained the same amount of *BTK* cDNA. The results of the two enrichment procedures also compare favourably; 104 of the 1536 subcloned cDNAs were found to map to the region contained in the YAC in the study by Vetrie *et al.* (approximately 7%) and in this study 1 out of 16 subclones mapped to the YAC (approximately 6%). Based on the above factors, there is very little difference between the methodology employed

by Vetrie et al. and the methods employed in this study.

It is apparent, however, that the success of Vetrie *et al.* lay in the ability to screen rapidly a large number of clones using automated griding facilities. The construction of a cosmid contig over the YAC then enabled the sub-cloned cDNAs to be quickly grouped. There was not much difference between the levels of enrichment obtained in the two studies, but the screening method used in this study relied on a much greater level of efficiency which had not been obtained. Vetrie *et al.* reported an enrichment of the *BTK* cDNA of 500 fold, compatible with the 300 fold enrichment of the *GLA* positive control.

In conclusion, the technique of cDNA enrichment appeared to be an efficient method for the identification of the gene responsible for XLA from the defined genomic area. The success appears to rely on the optimisation of the strategy for quenching repeat sequences and the characterisation of a large number of recombinant clones.

### 4.3.3 Implications of the isolation of the gene for XLA

The isolation of the gene responsible for XLA ended a long search by several laboratories around the world. The gene was first localised to the X chromosome in 1986 (Kwan *et al.* 1986; Mensink *et al.* 1986) and the positional cloning process took seven years to complete. The concurrent isolation of the same gene by a group using the candidate gene approach illustrates the effectiveness of these two different techniques. The isolation of the *BTK* gene will now allow mutation analysis of the gene in XLA patients, leading to a further understanding of the molecular basis of this disease. An understanding of the molecular basis of XLA may then allow further research into the mechanism of B cell development, as well as providing better disease diagnosis, carrier detection and pre-natal diagnosis for families affected by this disease.

# CHAPTER 5
# DETECTION OF MUTATIONS IN THE *BTK* GENE CAUSING XLA

## 5.1 Introduction

A broad study of pathological mutations in *BTK* should provide a range of complete and partial knockouts of protein function which can then be correlated with the patients immunological phenotype. This may allow important residues and domains in the Btk protein to be defined, furthering knowledge of non-receptor protein tyrosine kinases in general and, more specifically, the role of Btk in B cell development. The identification of mutations will also provide improved disease diagnosis and carrier detection in families affected by XLA, thereby improving the accuracy of the information upon which genetic counselling is based.

Only a small number of mutations had been documented prior to this study. Vetrie *et al.* (1993) identified genomic deletions in the *BTK* gene in five XLA patients, two of which were entirely intragenic, confirming the relationship between the *BTK* gene and XLA. Three patients with altered restriction patterns, which were co-inherited with XLA in the affected family, were also identified. Further analysis of two of these patients, who were found to have the gain and the loss of a *Taq*I restriction site, respectively, identified point mutations which would result in amino acid substitutions at residues thought to be crucial for the kinase function of non-receptor tyrosine kinases. Patient A was found to have a G to A transition resulting in the substitution of arginine 525 by glutamine, the change of a conserved amino acid in the substrate specific domain of the kinase. Patient B had an A to G transition causing the substitution of lysine 430 by glutamic acid in the ATP binding site of Btk. These mutations implicated *BTK* in XLA pathogenesis, and indicated that the kinase activity of the protein was important for Btk function.

The 659 amino acid Btk protein has been detected in other haematopoietic cells including myeloid cells, in addition to cells of the B lymphocyte lineage (de Weers *et al.* 1993; Genevier *et al.* 1994a; Smith *et al.* 1994a). Boys affected by XLA, however, have no evidence of myeloid abnormality and X inactivation patterns are non-random only in the B cells of carrier women, while being random in whole blood. This suggests that Btk is essential only for the development of B cells. At the time of this study, the genomic structure of this gene was not known, and therefore DNA could not be used as the basis of a mutation analysis study. As XLA boys have no B cells, the expression of the gene in myeloid cells allowed the study of the mutant gene in the form of RNA, and therefore cDNA, from purified mononuclear cells from their blood.

All patient DNAs available to us (approximately fifty families) were screened for deletions by Southern blot analysis using the *BTK* cDNA as a probe, prior to this more detailed mutation analysis study. Approximately 10% were shown to have genomic deletions (Lovering *et al.* 1994). The remaining 90% were therefore thought to have small insertions, deletions and point mutations. The definitive way to find such mutations is to completely sequence all exons, splice sites and control elements, but with a large gene and multiple samples a screening method is more appropriate. A mutation can then be located to within a small region of the gene, which must then be sequenced to define the mutation exactly. The screening method chosen for this study was single strand conformation polymorphism (SSCP) analysis (Orita *et al.* 1989a; Orita *et al.* 1989b). This technique is quick, simple and sensitive and does not require specialist chemicals or equipment.

## 5.2 Development of SSCP analysis for the *BTK* gene

### 5.2.1 The technique

SSCP analysis works on the principle that single stranded DNA, under non-denaturing conditions, will form a secondary structure which is sequence dependent. Analysis of denatured double stranded DNA on polyacrylamide non-denaturing gels allows the single

strands to migrate, not only according to size, but according to their sequence dependant secondary structure. Radioactive labelling or other staining techniques allow the single strands to be visualised. A single base change in a fragment up to 400bp can alter the secondary structure, resulting in a change in the pattern of migration. This is detected as a shift in one or both of the single strands which localises the sequence alteration to that defined area of the gene. This area must then be sequenced to define the exact nucleotide change.

SSCP has been used widely in mutation analysis studies (for example Hayward *et al.* 1994; Shimizu *et al.* 1994), and has been reported as a sensitive technique, detecting 80-90% of mutations in any particular gene (Michaud *et al.* 1992; Sheffield *et al.* 1993; Fan *et al.* 1993). As the technique relies on the formation of secondary structures sensitive to sequences changes, it does not, generally, identify 100% mutations and this must be borne in mind when undertaking studies of this type. The sensitivity of SSCP is dependent on the size of the fragment under analysis, approximately 150bp has been reported to be the optimum (Sheffield *et al.* 1993), and is inversely proportional to the length of the fragment, although strictness of this rule appears to vary depending on the sequence under analysis (Fan *et al.* 1993). Sensitivity of SSCP is highest when the mutation involves the insertion or deletion of base pairs from the sequence, but nucleotide substitutions can be identified almost as well. Transitions and transversions are identified at equal rates indicating that the type of base change involved does not change the sensitivity of detection (Sheffield *et al.* 1993; Fan *et al.* 1993). The position of the change in the fragment under analysis has been reported to be important in one study (Sheffield *et al.* 1993) but not in another (Fan *et al.* 1993). As this technique relies on the conformation formed by the single strands in the gel, it is sensitive to the physical environment of the gel; the temperature, buffer concentration, cross-linking of the acrylamide and concentration of acrylamide have all been shown to affect this type of analysis (Michaud *et al.* 1992; Sheffield *et al.* 1993; Fan *et al.* 1993). Low concentrations of glycerol have been shown to improve the separation of the strands (Orita *et al.* 1989b). Overall, room temperature gels with 5% glycerol and 0.5 x TBE appear to provide a good level of detection (Sheffield *et al.* 1993), and additional runs at 4°C with 10% glycerol or 0% glycerol and 1 x TBE or 0.5 x TBE provide small

increases in sensitivity (Michaud *et al.* 1992).

### 5.2.2 Application of SSCP to analysis of *BTK*

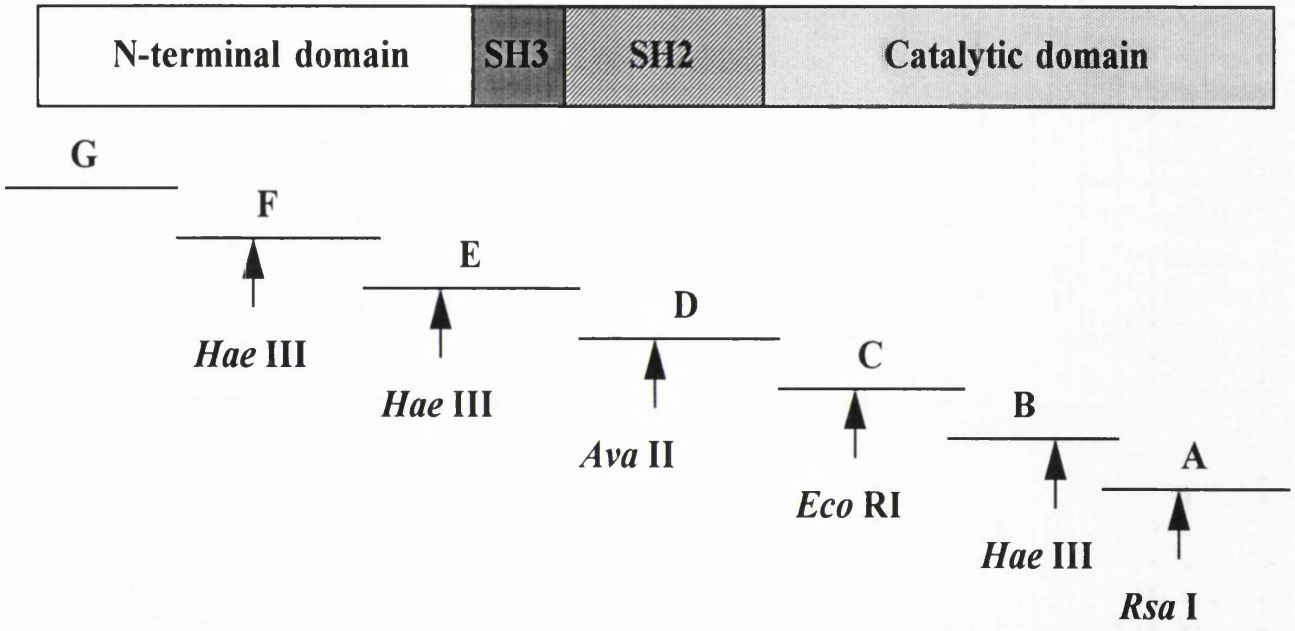
The protein coding region of the *BTK* cDNA covers nearly 2kb. PCR primers were designed which allowed the amplification of the protein coding region of the cDNA in seven overlapping sections, each of approximately 400bp. The primers were all designed to have approximately the same annealing temperature to allow all combinations of primers to be used in reactions together. For each fragment, restriction enzyme sites were selected which allowed the digestion of the product into two smaller fragments, as shown diagrammatically in Figure 5.1 and as described in Table 2.1. The sites were selected so that each of the fragments produced were greater than 100bp in length and that the difference in length between the two fragments was at least 50bp, to allow the single strands to be separated unambiguously on the gel. All samples were analysed at room temperature and at 4°C to maximise the sensitivity of the technique. Any difference between patient and control was scored as positive. Undenatured and denatured double stranded DNA from a normal control was analysed on every gel to allow identification of the normal single strands. In all cases, if a band shift was detected in a particular fragment, all the other reactions across the cDNA were performed and found to be normal, indicating that there was only one mutation in each patient.

## 5.3 Results

### 5.3.1 Mutations detected in SSCP reaction A

The PCR product from reaction A was digested into a 195bp fragment and a 119bp fragment. Four patients from three families were found to have band shifts in the fragments from this reaction on SSCP analysis. Patient AP had a band shift in the most 3' fragment of the cDNA, the smaller 119bp fragment (Figure 5.2a). DNA sequence analysis of this fragment showed that there had been a mutation of a G to a T at nucleotide position 2038 (Figure 5.2b), resulting in a premature stop codon at amino acid residue 636.

**Figure 5.1 Diagrammatic representation of SSCP Analysis for the Btk gene**




A-G are PCR amplification products as detailed in section 2.12.1. Arrows indicate approximate position of restriction enzyme digest sites.

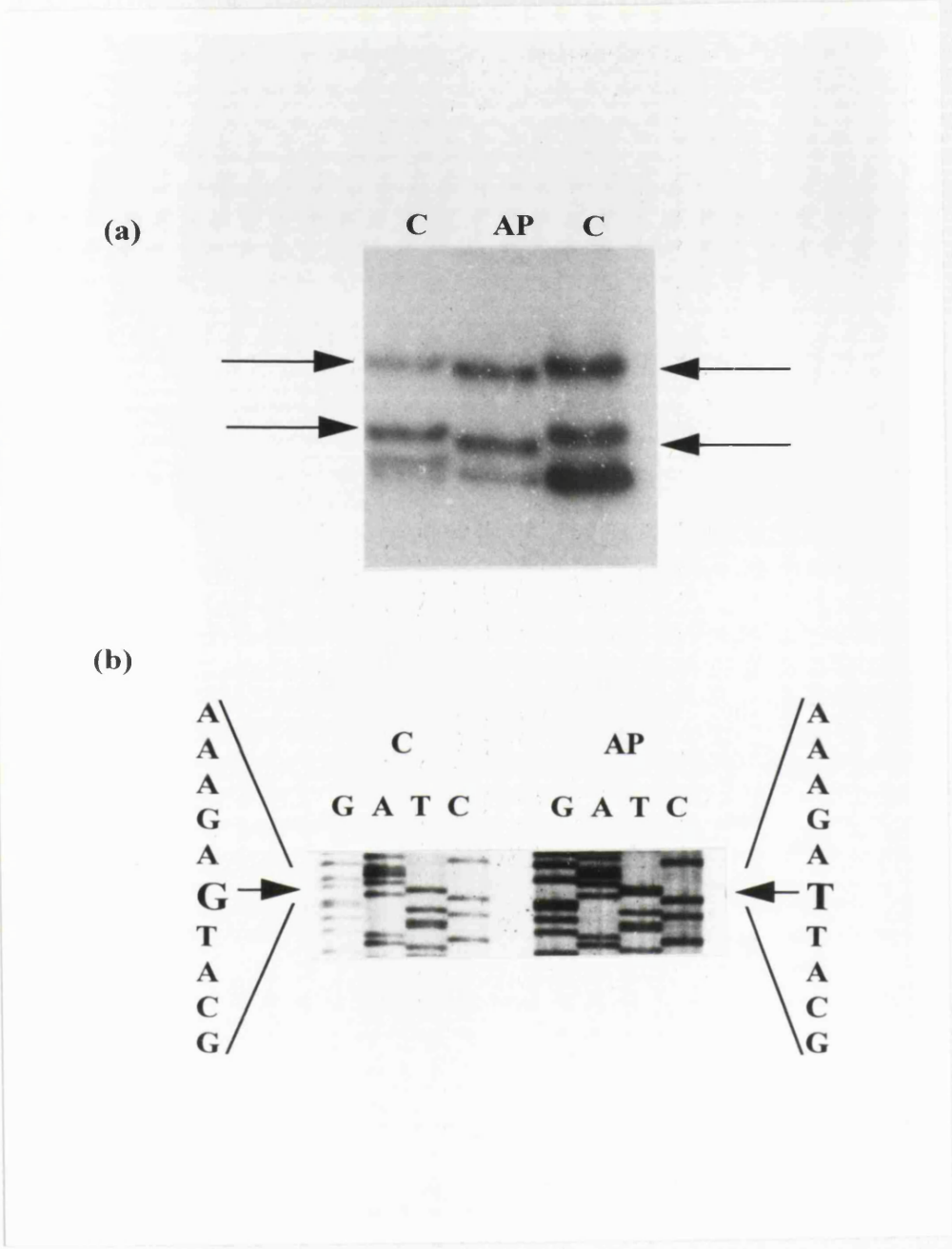
Three XLA patients from two families had an SSCP band shift in the 195bp fragment from reaction A. The three affected boys from Family F were initially diagnosed as having an undescribed B cell immunodeficiency but on the basis of a non-random pattern of X-inactivation in the B cells of the obligate carrier female (Alterman *et al.* 1993), they subsequently were diagnosed as having a "leaky" form of XLA. The three boys in this family have B cell counts up to 2%. Immunoglobulin levels vary considerably between boys, but are almost normal. Clinical details are given in Appendix II. These boys are not on intravenous immunoglobulin therapy. Brothers TF and JF were studied by SSCP analysis and found to have a shift in the larger fragment of reaction A (Figure 5.3a). This band shift was shown, on sequence analysis, to be the result of a C to A mutation at nucleotide position 1952 (Figure 5.3b), which would lead to the substitution of alanine by aspartic acid at amino acid position 607.

The band shift seen in the *BTK* cDNA in patient AJ (Figure 5.4a) was found, on DNA sequence analysis, to be caused by the insertion of two nucleotides, CA or AC, at nucleotide position 2013 or 2014 of the *BTK* gene (Figure 5.4b). This dinucleotide insertion would result in the *BTK* gene in this patient encoding an out of frame protein from threonine 628 onwards, with a premature termination codon at amino acid 649.

### 5.3.2 Mutations detected in SSCP reaction B

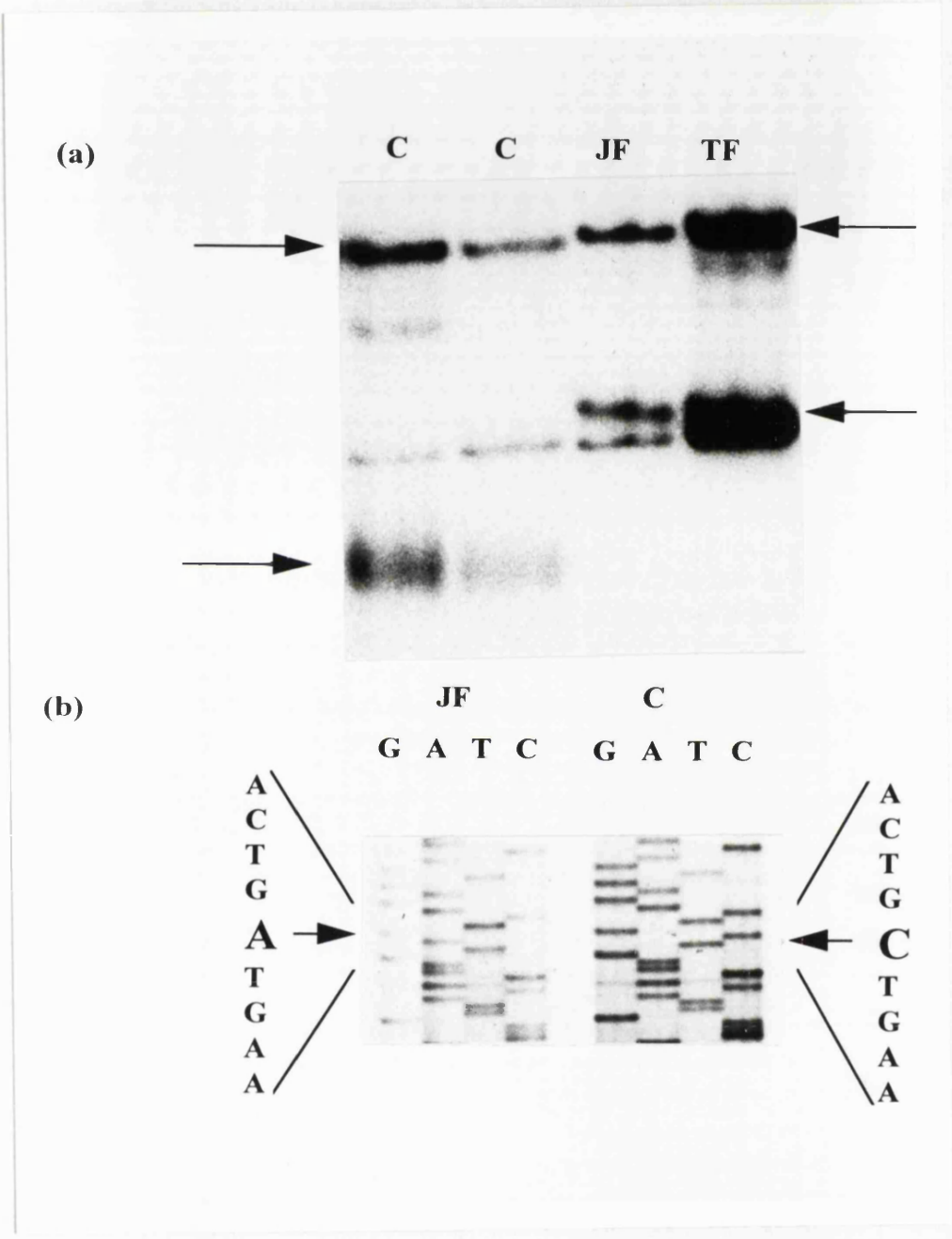
The larger fragment (168bp) produced by the digestion of the product from reaction B was shifted in patient FM (Figure 5.5a). This shift was only detected on the room temperature gels. The band shift in patient FM was shown to be caused by a G to A mutation at nucleotide position 1691 (Figure 5.5b), resulting in the substitution of arginine 520 by glutamine.

**Figure 5.2 Detection of a *BTK* mutation in patient AP**



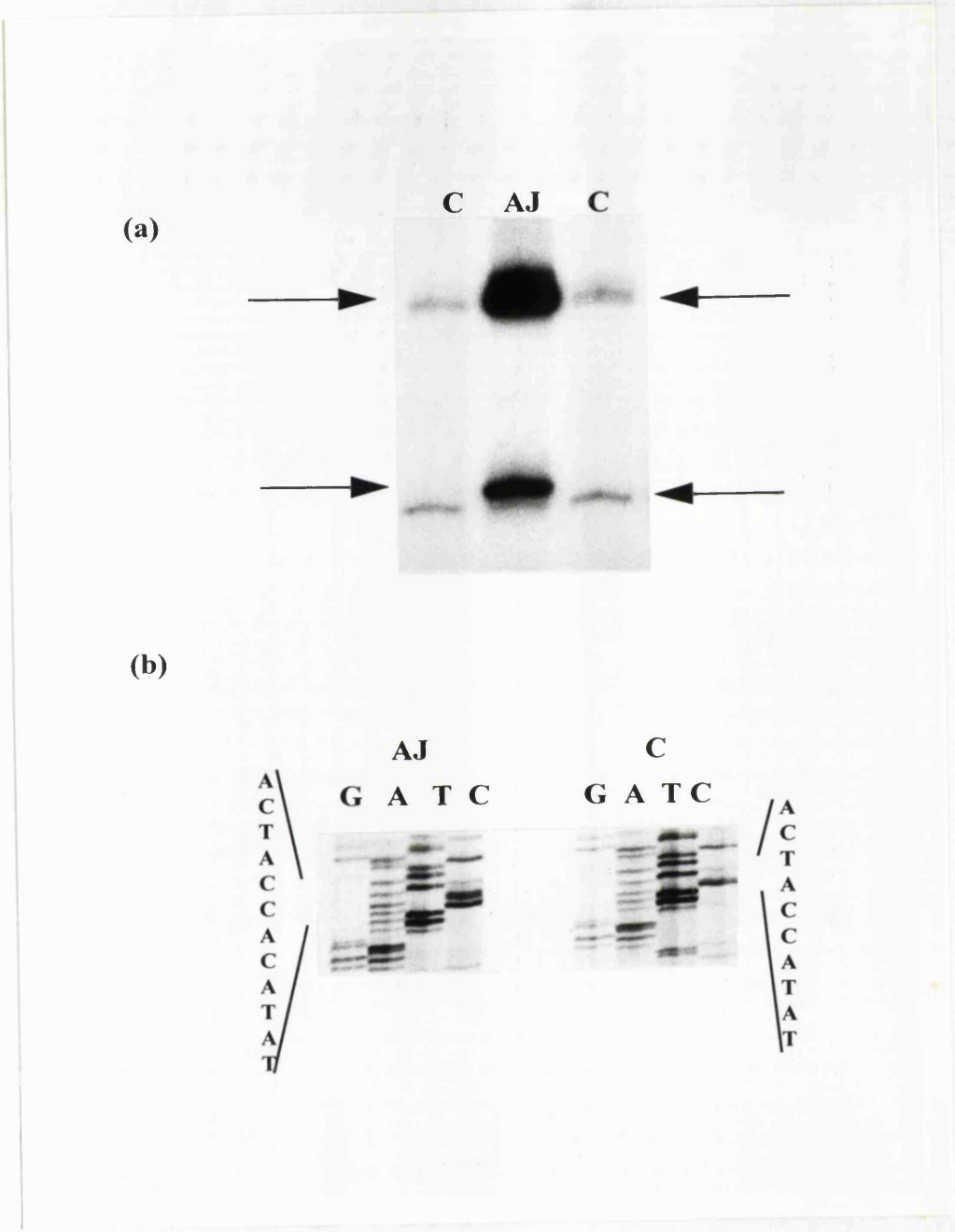

Mutation in patient AP compared to normal control (C) shown as (a) an SSCP band shift in the 119bp fragment from reaction A, with arrows on the left indicating the normal SSCP pattern and arrows on the right indicating the mutant pattern, with the lowest of the three bands being the non-denatured double stranded fragment, and (b) an autoradiograph of a DNA sequencing gel showing the G to T transversion at position 2038 in the *BTK* gene.

**Figure 5.3 Detection of *BTK* mutation in family F**



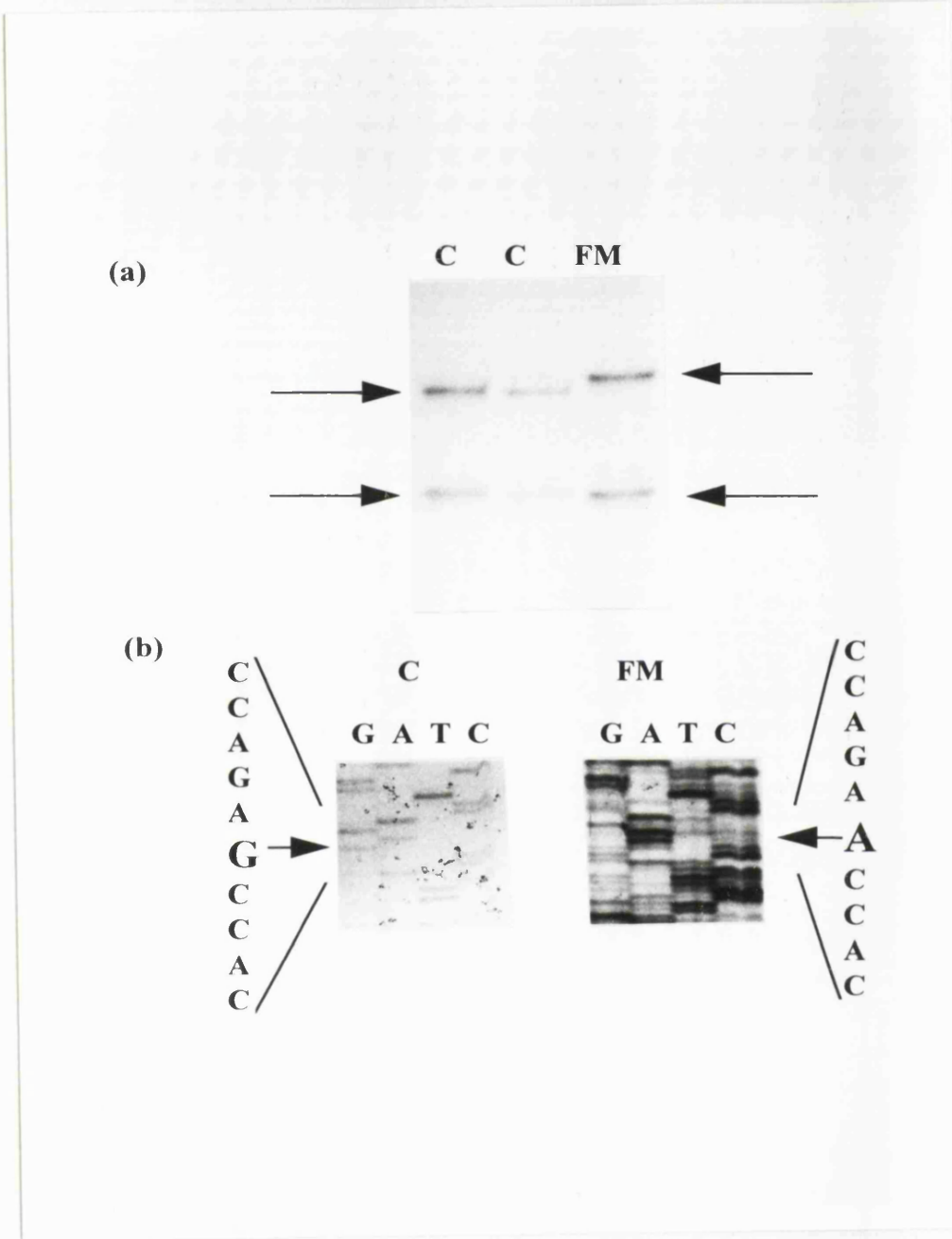

Detection of a mutation in family F (a) by SSCP analysis of brothers JF and TF showing a band shift in the 195bp fragment from reaction A (arrows on the right) compared to the band pattern seen in two normal controls (C) (arrows on the left) and (b) by DNA sequence analysis of JF, showing the C to A transversion at position 1952 of the *BTK* gene.

**Figure 5.4 Detection of the mutation in patient AJ**




Analysis of the *BTK* geme from patient AJ showing (a) an SSCP band shift in the 195bp fragment from reaction A, with the aberrant pattern indicated on the right and the normal pattern from the controls (C) indicated on the left and (b) the insertiom of a dinucleotide (CA/AC) at position 2013/2014 of the *BTK* gene.

**Figure 5.5 Mutation analysis of patient FM**




(a) an SSCP band shift in the 168bp fragment of reaction B (indicated on the right) compared to a normal control (C) (indicated on the left) and (b) a G to A transition at position 1691 of the *BTK* gene identified by DNA sequence analysis.

### 5.3.3 Mutations detected in SSCP reaction C

No patients studied showed mutations in either of the fragments generated from this reaction.

### 5.3.4 Mutations detected in SSCP reaction D

Three patients showed SSCP band shifts in the smaller (144bp) fragment produced from reaction D. DNA sequence analysis of this fragment from patient JP showed that the band shift (Figure 5.6a) was caused by an A to G mutation at nucleotide position 1051 (Figure 5.6b), which would result in arginine 307 being substituted by glycine.

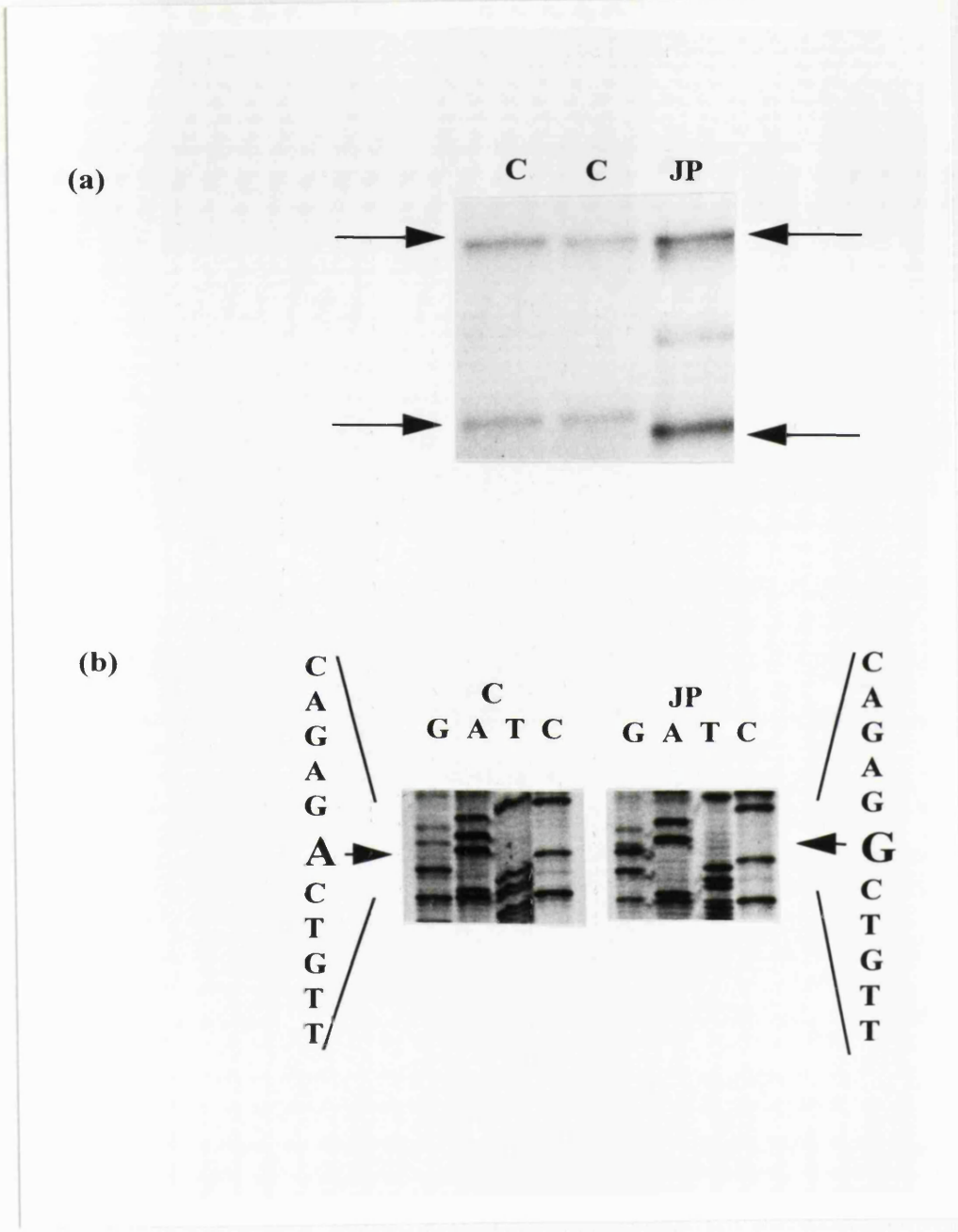
Patient 276 is a member of a large XLA pedigree whose members exhibit heterogeneity in their clinical and immunological phenotypes (Mensink *et al.* 1984; de Weers *et al.* 1994b). SSCP analysis showed band shifts in the smaller fragment (144bp) from reaction D (results not shown) and the larger fragment (226bp) from reaction E (Figure 5.7a), suggesting that the nucleotide change lay in the region where these two fragments overlap (nucleotides 988-1000). Sequence analysis showed that the shift was due to a C to T mutation at nucleotide position 994 (Figure 5.7b). This base substitution results in arginine 288 being replaced by tryptophan.

The third patient, AC, showing a band shift in the smaller fragment (144bp) from reaction D (Figure 5.8a), caused by a trinucleotide deletion (Figure 5.8b). The deleted AGG could either be nucleotides 1035 - 1037 or nucleotides 1038 - 1040, but either of these deletions would result in the deletion of the glycine residue at position 302.

### 5.3.5 Mutations detected in SSCP reaction E

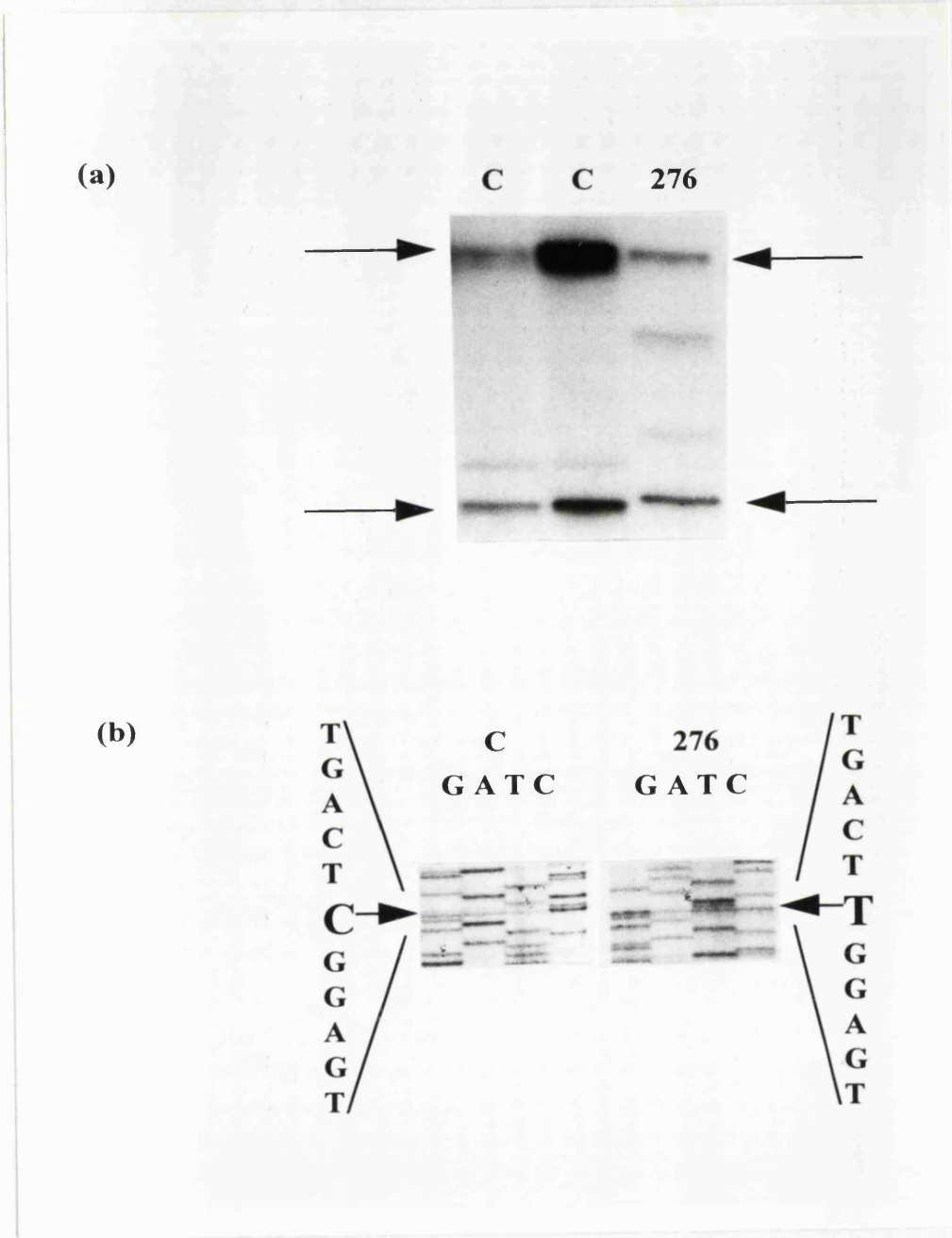
In addition to patient 276 (see section 5.3.4), the larger (226bp) fragment of reaction E was also shifted in patient NAD (Figure 5.9a). Sequence analysis showed a C to T mutation at nucleotide position 895 (Figure 5.9b) which would result in a premature stop codon at amino acid residue 255.

**Figure 5.6 Detection of a mutation in the *BTK* gene from patient JP**



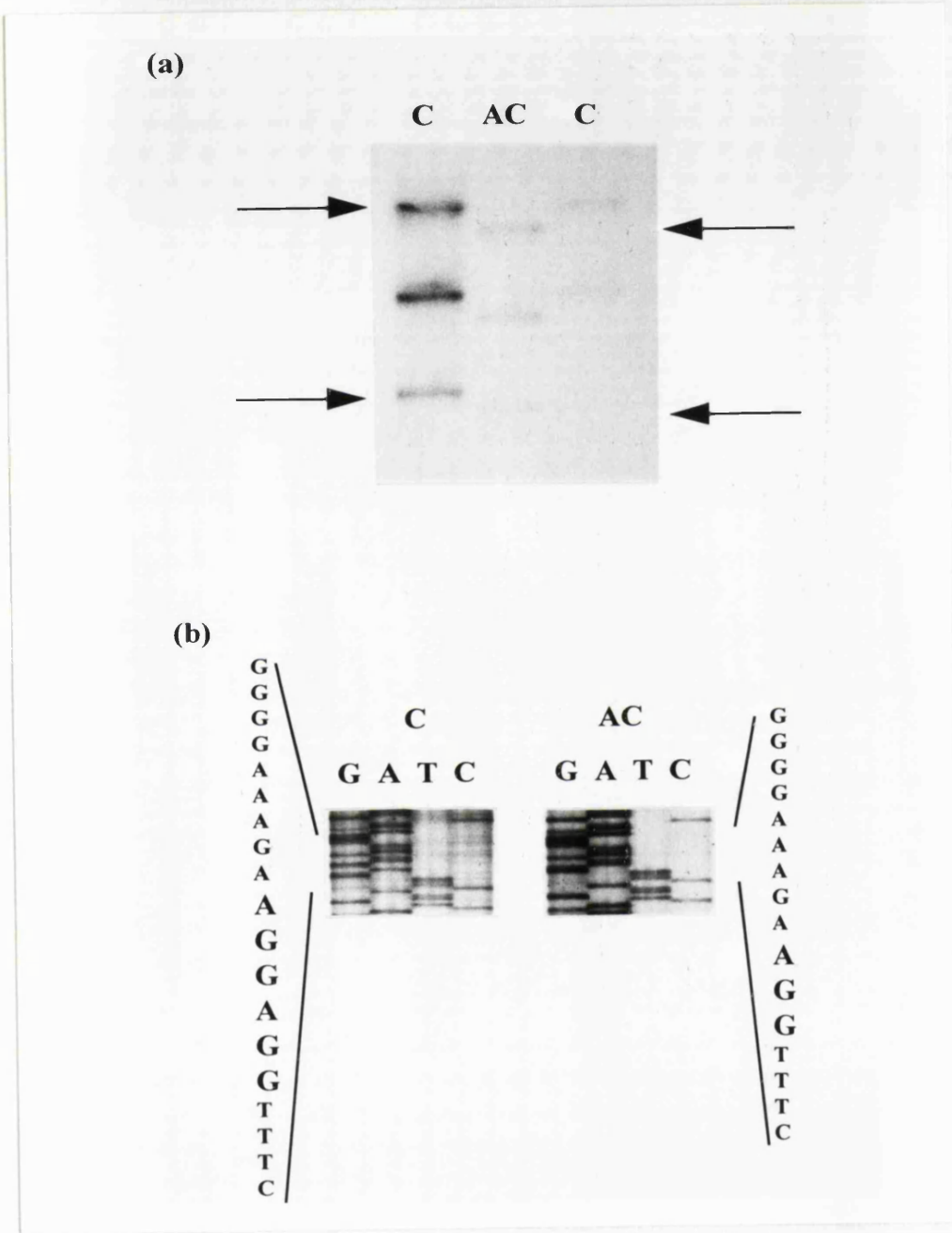

(a) SSCP band shift in the 144bp fragment from reaction D - patient JP is shown on the right and the normal controls (C) are shown on the left, with arrows indicating the band patterns and (b) DNA sequence analysis of the same fragment which revealed an A to G transition at position 1051 in the patient, compared to the normal control (C).

**Figure 5.7 Detection of a *BTK* mutation in patient 276**



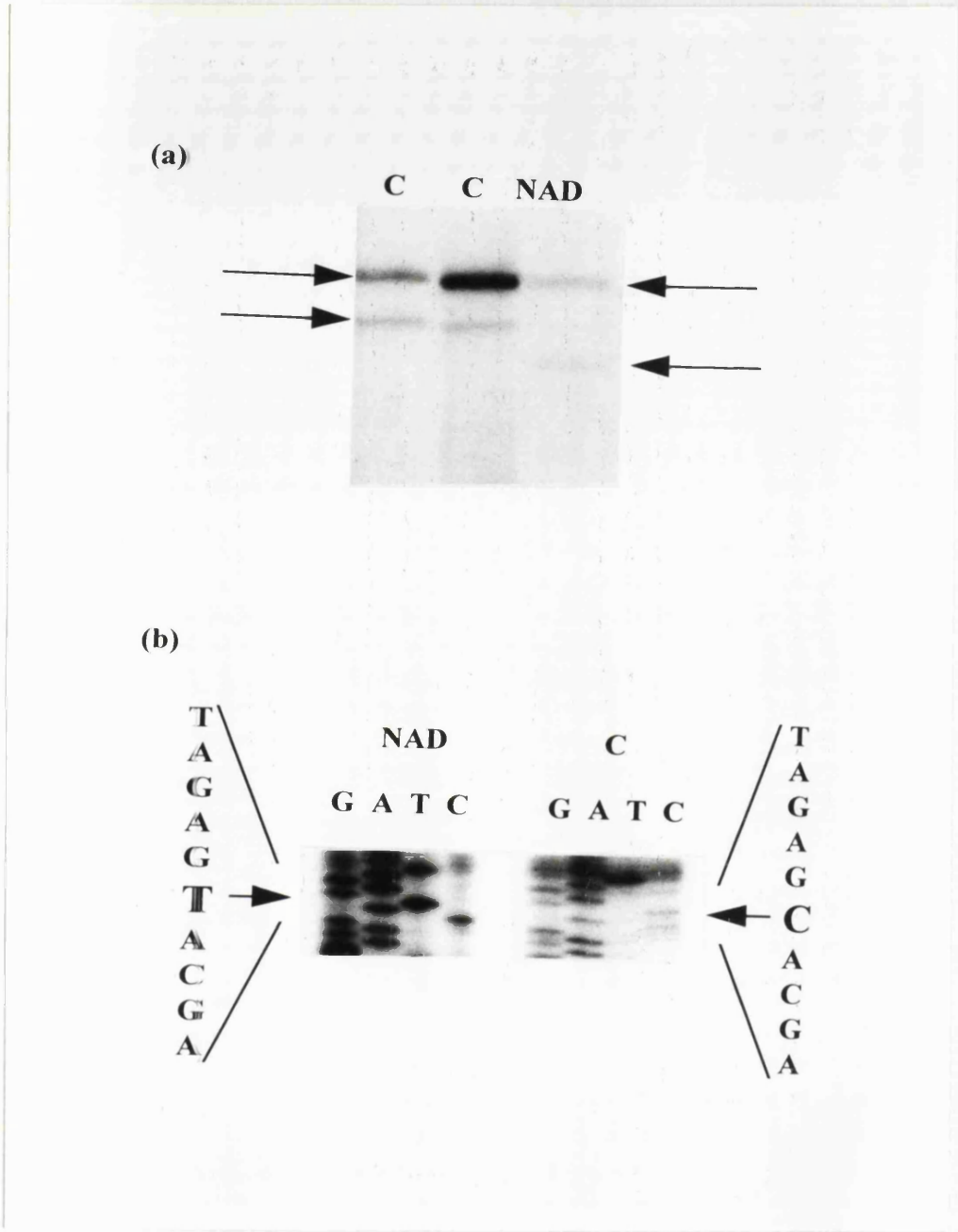

(a) SSCP analysis, showing a band shift in the 226bp fragment from reaction E (a band shift was also seen in the small fragment (144bp) of reaction D (results not shown), compared to two normal controls (C), with the band patterns indicated by arrows, and (b) DNA sequence analysis, showing the C to T mutation at position 994 of the *BTK* gene, compared to the normal control (C).

**Figure 5.8 Analysis of *BTK* in patient AC**




(a) an SSCP band shift in the smaller fragment (144bp) from reaction D in patient AC indicated by arrows on the right, compared to the normal control (C) indicated by arrows on the left and (b) the deletion of the trinucleotide AGG, at positions 1035-1037 or 1038-1040 in patient AC compared to the normal control (C), on DNA sequence analysis.

**Figure 5.9 Mutation detection in patient NAD**




(a) SSCP analysis of the *BTK* gene from patient NAD, showing an altered banding pattern in the 226bp fragment from reaction E compared to the normal control (C) (banding patterns indicated by the arrows on left and right respectively) and (b)[$^{32}$]P-dATP DNA sequence analysis showing the C to T mutation found at position 895 of the *BTK* gene

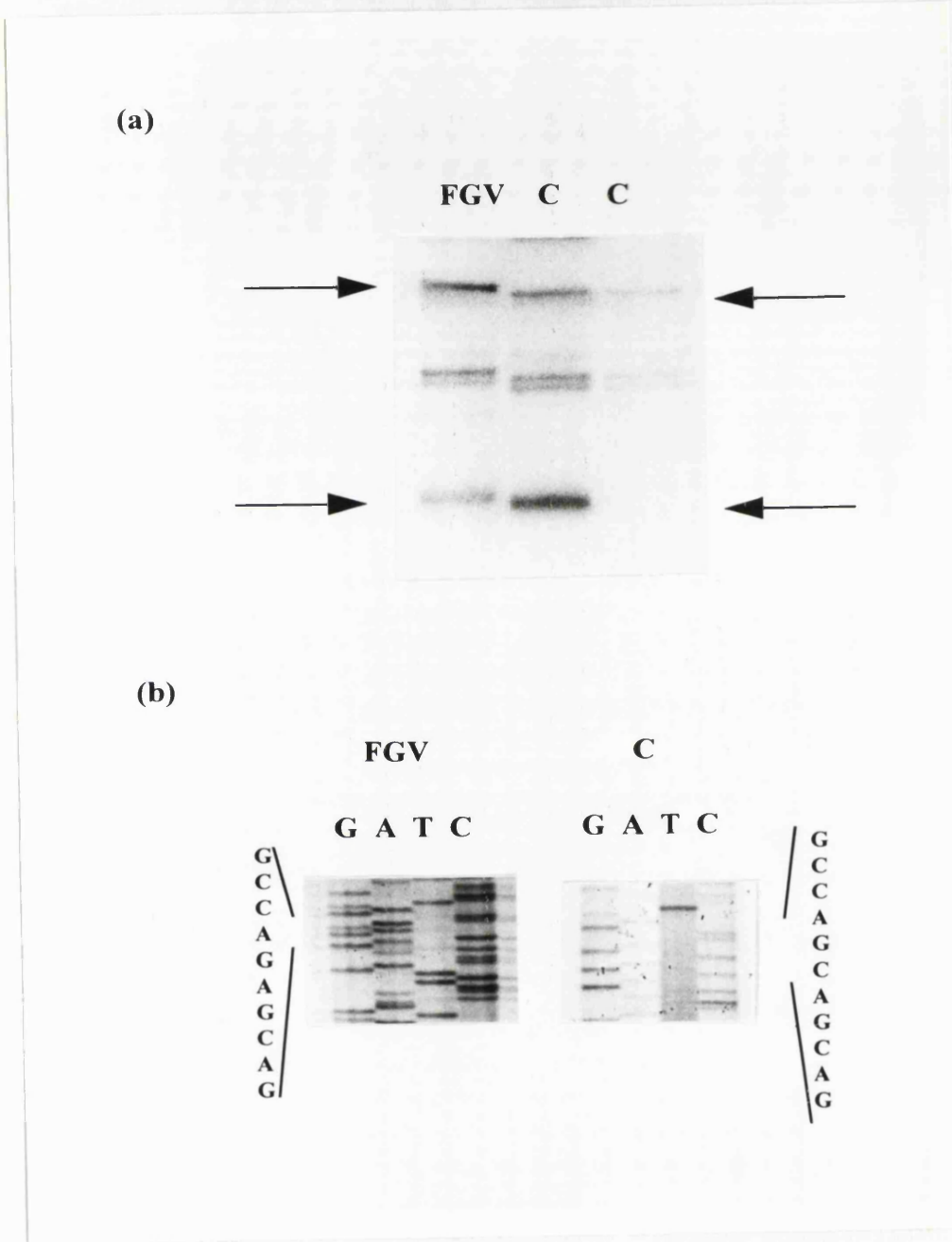
SSCP analysis of the *BTK* cDNA from patient FGV showed a band shift in the smaller fragment (175bp) of reaction E (Figure 5.10a), resulting from the deletion of a C at nucleotide position 752 (Figure 5.10b). The protein encoded by this mutant gene would be out of frame from alanine 207 and prematurely terminated at residue 216.

### 5.3.6 Mutations detected in SSCP reaction F

Patient JG had been studied previously by Southern blot analysis (patient C (Vetrie *et al.* 1993c)), and was shown to have an extra *Msp*I site in the N terminal region of the gene, resulting in an abnormal restriction fragment pattern. This patient has two affected brothers. SSCP analysis of the *BTK* gene from patient JG and his brother COB showed band shifts in the smaller (141bp) fragment from reaction F. Sequence analysis of the shifted band from fragment F in patient JG (Figure 5.11a) showed a 21 bp insertion (TCTGTGTTTTCATCGACCCGG) at nucleotide position 442 (amino acid position 103) (Figure 5.11b). This insertion would result in seven amino acids being introduced between amino acids 103 and 104; Q103-SVFSSTR-V104. Comparison with recently published *BTK* intronic sequences found immediately 5' to exon 5 (Hageman *et al.* 1994; Ohta *et al.* 1994; Sideras *et al.* 1994) showed that the inserted sequence was found in the genomic DNA and that this mutation must be caused, therefore, by a splicing error. Analysis indicated that there had been an A to G mutation at the -2 position of the intron/exon boundary at the 5'end of exon 5, destroying the conserved AG splice acceptor sequence. This mutation would also create the *Msp*I site discussed in Vetrie *et al.* (1993). A cryptic splice site 21bp upstream in the intron appears to be utilised.

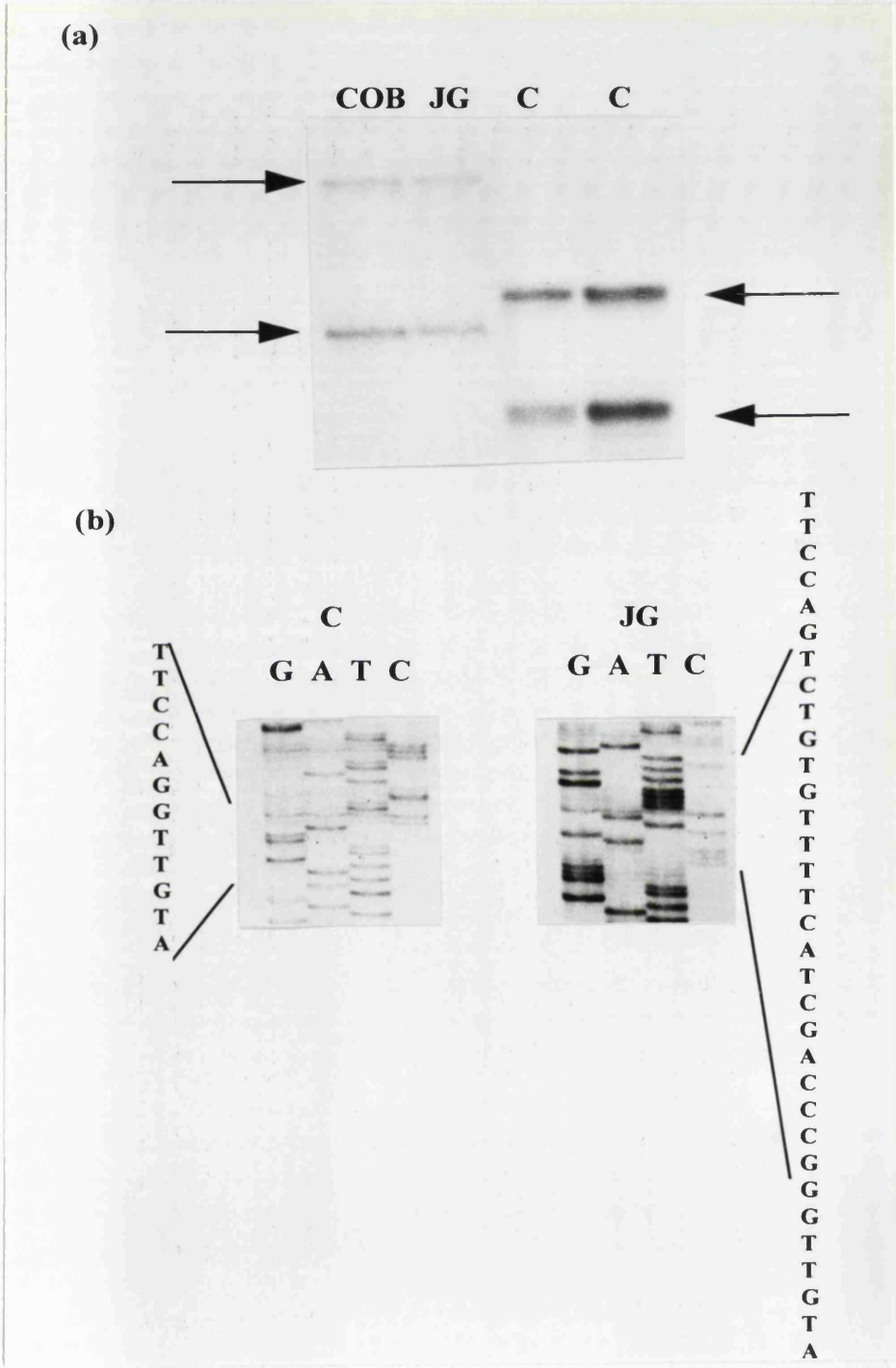
Patient LT was also found to have a band shift in the smaller fragment (141bp) of reaction F (Figure 5.12a). This patient also had a band shift in reaction G (results not shown) which indicated that the mutation lay in the overlapping regions of reactions F and G (nucleotides 339-357). On sequence analysis, an additional A was found in a run of seven A residues, at nucleotides 341-347, within the overlap (Figure 5.12b). This mutation will result in the Btk protein being translated out of frame from residue 72 onwards, with premature termination at residue 83.

## Figure 5.10 *BTK* analysis in patient FGV



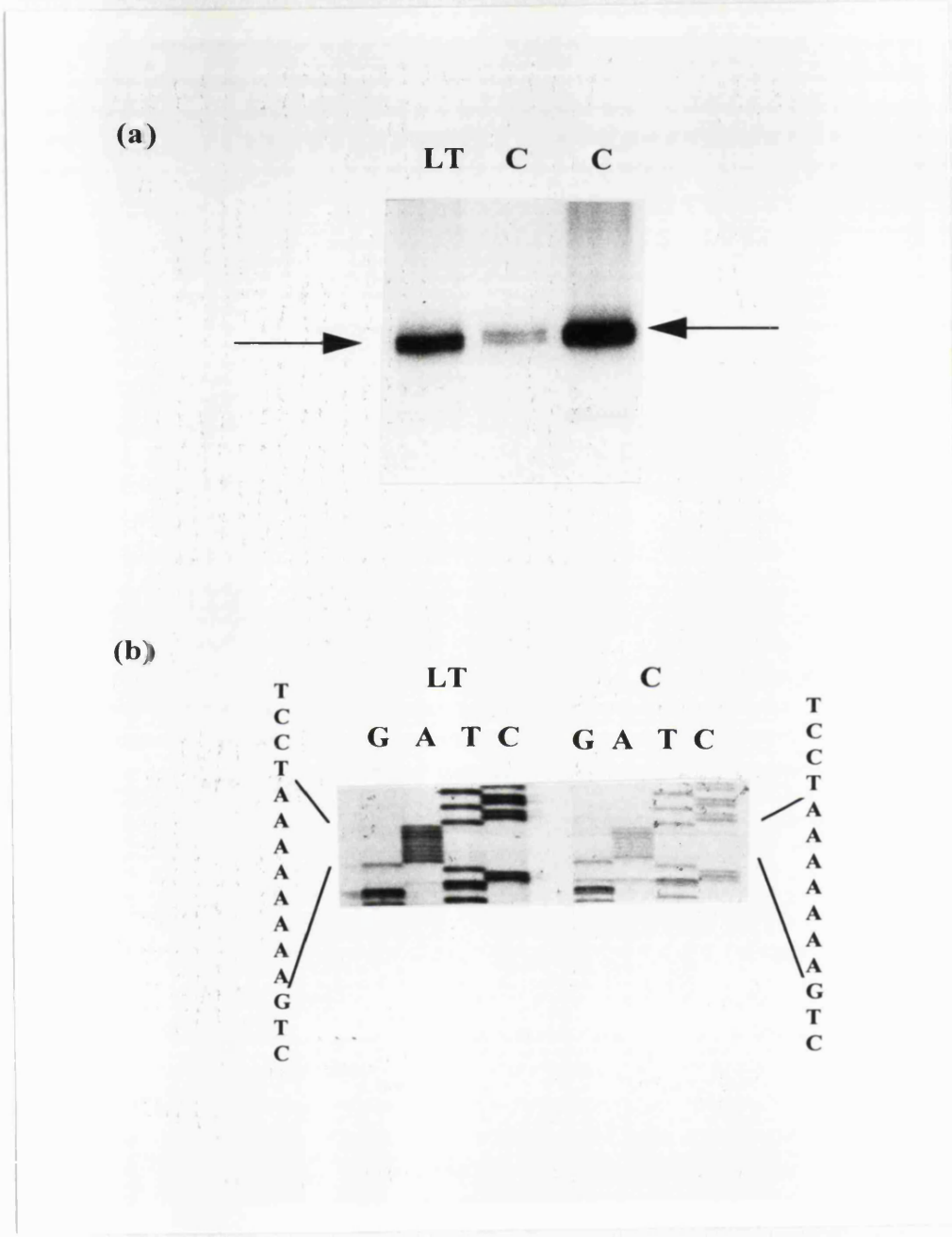

Detection of a mutation in patient FGV by (a) SSCP analysis, which showed a band shift in the 175bp fragment of reaction E in the patient when compared to the normal control (C), the banding patterns being indicated by arrows on the right and left, respectively, and (b) DNA sequence analysis, which showed deletion of a C nucleotide at position 752 of the gene.

**Figure 5.11 Mutation detection in patient JG**




(a) SSCP analysis of the *BTK* gene from patient JG and his affected brother COB showing a band shift in the 141bp band from reaction F, when compared to the normal controls (C), with normal and mutant band patterns indicated by arrows on the left and right, respectively, and (b) DNA sequence analysis of the same fragment showing a 21bp insertion at nucleotide 442.

**Figure 5.12 Analysis of *BTK* from patient LT**




(a) SSCP analysis of the 141bp fragment from reaction F, showing the band shift in patient LT as compared to the normal control (C), with normal and mutant band patterns shown on the left and right, respectively, and (b) DNA sequence analysis showing the insertion of an A in the monotonic run at nucleotides 341-347.

### 5.3.7 Mutations detected in SSCP reaction G

In addition to patient LT (see section 5.3.6), SSCP analysis of reaction G in the XLA patients showed band shifts in three patients from two families - SB, his brother CB and unrelated BB (Figure 5.13a). Sequence analysis of the cDNA from patients SB and BB showed that they both have mutations in the initiation codon, SB having an A to G mutation at position 133 and BB having a T to C mutation at position 134 (Figure 5.13b). This mutation probably results in the complete prevention of translation, unless an alternative downstream translation start site is used.

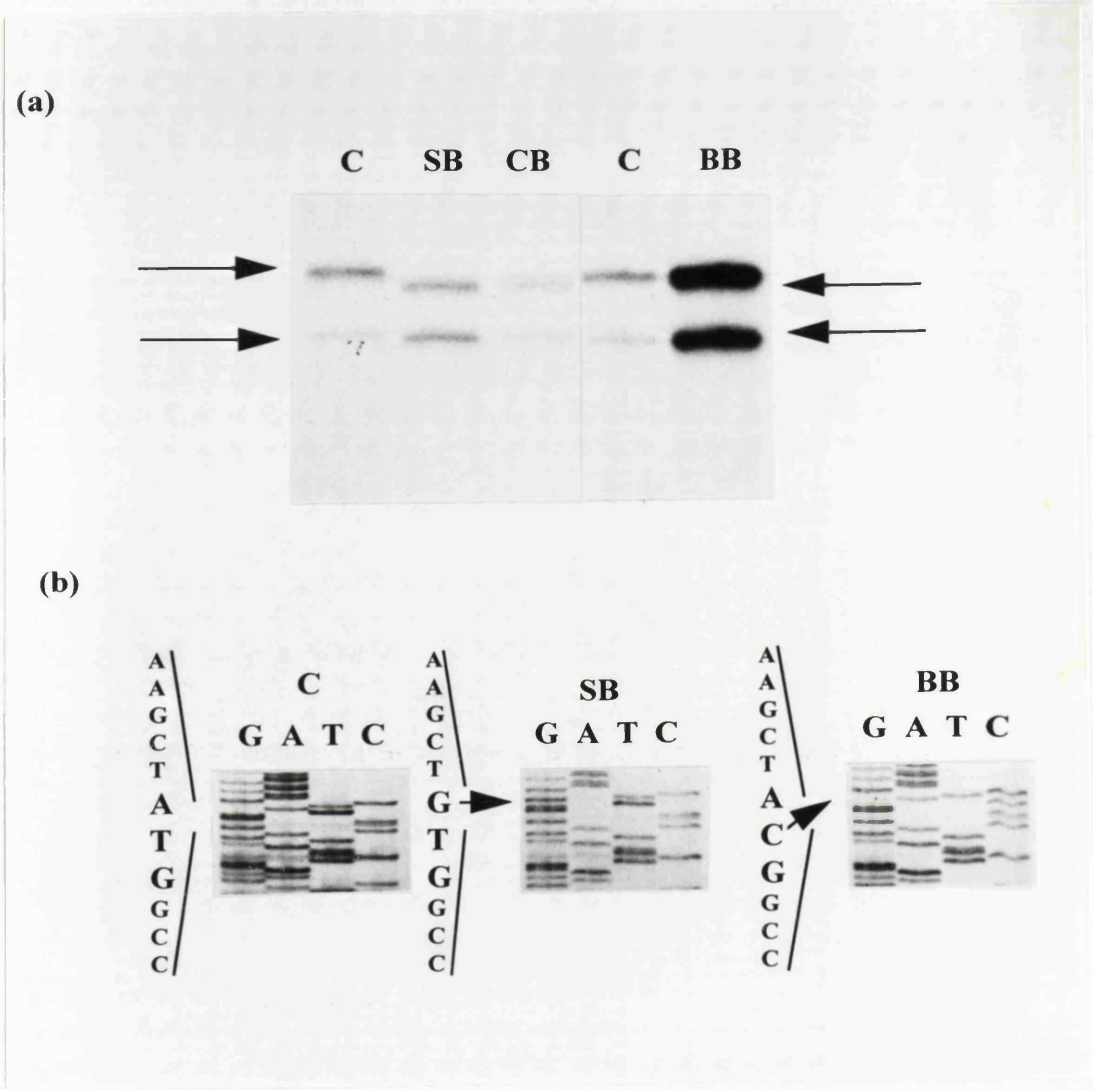
### 5.3.8 XLA patients with no alterations detectable by SSCP analysis

Eighteen XLA patients were analysed for mutation in *BTK* in this study. Three of these patients were fully analysed by SSCP and no alterations were found.

### 5.3.9 Two patients with exonic deletions

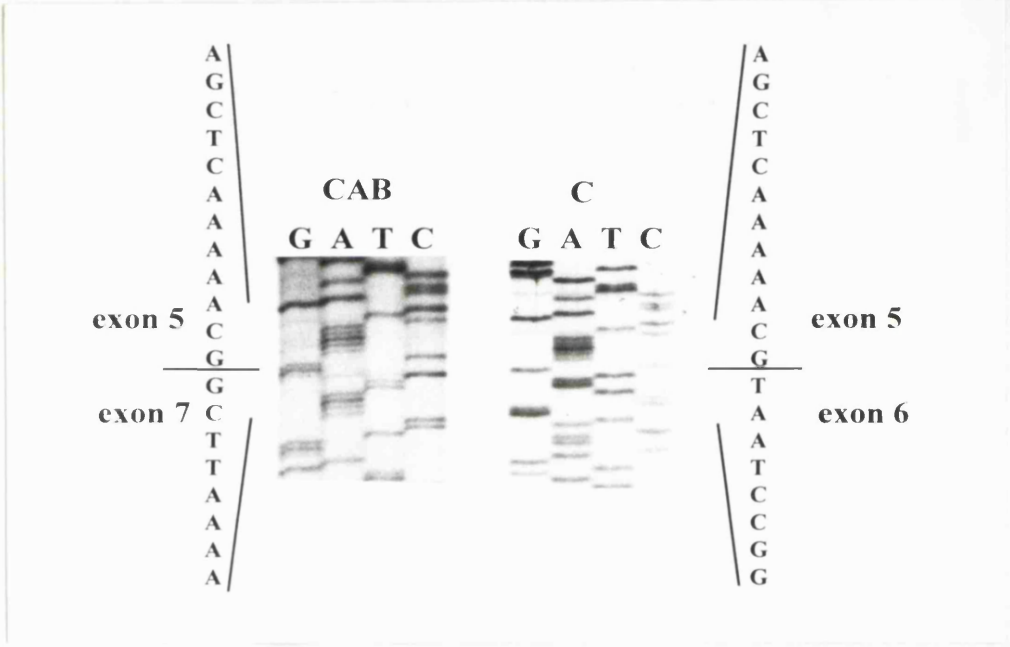
Prior to SSCP analysis, all patient cDNA samples were analysed by PCR amplification, to assess the presence on the normal sized *BTK* gene. Two patients showed smaller product sizes, indicative of a deletion in the cDNA (results not shown). These patients were analysed by DNA sequencing directly. Patient CAB, from a large Spanish pedigree, showed a 1kb PCR product using PCR primers StartF and 7R compared to the expected product size of 1.2kb. Southern blot analysis of DNA from affected and unaffected members of this family using a fragment from the 5' end of the *BTK* gene as a probe showed no alterations, suggesting that the deletion results from a splicing defect (S.Genet, personal communication). DNA sequence analysis of the region spanning the deletion showed that this patient had a deletion encompassing nucleotides 524-652 inclusive (Figure 5.14). Comparison with the recently published genomic structure of *BTK* indicated that these nucleotides corresponded to exon 6 (Hageman *et al.* 1994; Ohta *et al.* 1994; Sideras *et al.* 1994). Sequence analysis of the intron/exon boundaries should define the exact mutation producing this splicing error. The removal of this exon leaves the remainder of the protein being translated in the correct reading frame, but 43

**Figure 5.13 Mutation analysis of patients BB and SB**



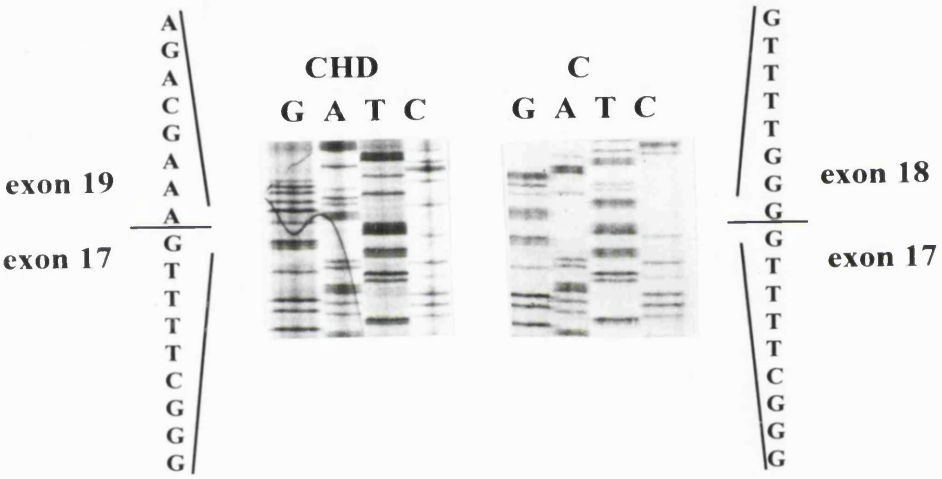

(a) SSCP analysis of fragment G from brothers SB and CB, unrelated BB and normal control (C), showing altered banding patterns indicated by arrows and (b) DNA sequence analysis showing A to G transition at position 133 and T to C transition at position 134 in patients SB and BB, respectively

**Figure 5.14 DNA sequence analysis of the *BTK* gene in patient CAB**




DNA sequence analysis showing the deletion of nucleotides 524-652 (corresponding to exon 6) from the *BTK* cDNA in patient CAB compared to a normal control (C).

**Figure 5.15 DNA sequence analysis of the *BTK* gene from patient CHD**




DNA sequence analysis showing the deletion of nucleotides 1883-2040 (corresponding to exon 18) from the ***BTK*** cDNA in patient CHD, compared to the normal control (C).

amino acids will be deleted and amino acid 131 will be changed to glycine. From genetic linkage studies of this family, the origin of the mutation is thought to have occurred in the unaffected grandfather's germ-line (S.Genet, personal communication).

Patient CHD also showed abnormalities on initial PCR analysis. PCR amplification using primers E3R and BS5' showed a product of approximately 350bp, compared to the expected product of 500bp. Sequence analysis of the patients cDNA showed that nucleotides 1883-2040 (inclusive) were missing (Figure 5.15). Comparison with the *BTK* genomic structure indicated that this corresponded to the absence of exon 18, suggesting the presence of a splicing defect. This error would result in the Btk protein being translated out of frame from amino acid 585 onwards, with premature termination at residue 588.

## 5.4 Discussion

### 5.4.1 Efficiency of *BTK* SSCP analysis

In this study, eighteen XLA patients were investigated for mutations in the *BTK* cDNA using SSCP analysis and the results are summarised in Table 5.1. Two of these patients were found to have deletions of whole exons indicating splicing errors, and thirteen were found to have SSCP band shifts caused by sequence alterations. This gives a detection efficiency for SSCP analysis of over 80%, which is comparable with detection rates reported in other studies (Michaud *et al.* 1992; Sheffield *et al.* 1993). Three patients did not have mutations in the *BTK* cDNA which were detectable by this method. These patients may have alterations in untranslated regions or control elements. Alternatively, as two of them are sporadic cases and X-inactivation analysis has not been performed to confirm X-linked inheritance, it is possible that their diagnoses as XLA are incorrect.

Nearly all the band shifts were detected on gels at both 4°C and at room temperature, but one (patient FM, section 5.3.2) was detected at room temperature only. Some fragments

## Table 5.1 Summary of XLA mutations

| Patient | XLA phenotype[1] | Fragment with SSCP band shift[2] | Mutation | Exon | Restriction digest site changes[3] | Protein | Comments |
|---|---|---|---|---|---|---|---|
| BB | classical | G | T134 → C | 2 | *EagI* + *Bsi*EI + | M1T | disruption of initiation codon |
| SB | classical | G | A133 → G | 2 | none | M1V | disruption of initiation codon |
| LT | classical | F/G | insertion A 341-347 | 3 | none | N72 | out of frame |
| JG | classical | F | insertion 21bp A→G -2, 3'splice site | 4/5 | *Msp*I+ | Q103-SVFSSTR | 7 amino acid insertion |
| CAB | classical | n/d | deletion 524-652 splicing error | 6 | n/d | V131G + del 43 amino acids | splicing error in-frame deletion |
| FGV | classical | E | deletion C 752 | 8 | *Cac*8I- | A207 | out of frame |
| NAD | classical | E | C895 → T | 8 | *Nla*III + | R255★ | premature stop |
| 276 | leaky | D/E | C994 → T | 10 | *Mly*I- | R288W | substitution SH2 domain |
| AC | classical | D | deletion AGG 1035 - 1037 or 1038-1040 | 11 | none | deletion G302 | single amino acid deletion SH2 domain |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| JP | classical | D | A1051 → G | 11 | none | R307G | substitution SH2 domain |
| FM | classical | B | G1691 → A | 15 | *Plf*MI + | R520Q | substitution kinase domain |
| TF | leaky | A | C1952 → A | 18 | *Bsm*I + | A607D | substitution kinase domain |
| AJ | classical | A | insertion CA/AC 2013/2014 | 18 | *MslI* + | T628 | out of frame |
| CHD | classical | n/d | deletion 1883-2040 splicing error | 18 | n/d | Y545 | splicing error out of frame |
| AP | classical | A | G2038 → T | 19 | *NlaIII* - | E636★ | premature stop |

[1]The classical XLA phenotype is defined as the patient having less than 1% B cells and undetectable immunoglobulin and the leaky XLA phenotype is defined as the patient having more than 1% B cells and some immunoglobulin. [2]SSCP reaction nomenclature as in Figure 5.1. [3] Restriction digest site changes in genomic DNA. + and - indicate gain and loss of a restriction site, respectively, included here for completion. Nucleotide and amino acid residue numbers are as in Vetrie *et al.* (1993) Exon nomenclature as in (Hageman *et al.* 1994; Ohta *et al.* 1994). n/d indicates not determined. ★ indicates a stop codon.

gave bands on SSCP analysis which were not sharply resolved under certain conditions, for example one of the single strands produced on SSCP analysis of the 195bp fragment of reaction A was not sharply resolved on the gel run at 4°C, but was much clearer under analysis at room temperature (see Figure 5.3a). Other fragments, for example the 226bp fragment from reaction E, produced multiple bands under certain conditions and in some cases the non-denatured double stranded DNA showed abnormal migration (see Figure 5.2a). These phenomena have been noted previously in other studies (Michaud *et al.* 1992). Many of the band shifts were more noticeable at 4°C, suggesting that these conditions were more sensitive, and it was found that the 4°C conditions were more reproducible, probably because of the more constant temperature. SSCP analysis at one temperature is therefore sufficient to give a high level of sensitivity, and it may be more efficient to only use a second set of conditions when no shifts are seen or the shift is very small and requires confirmation.

### 5.4.2 CpG dinucleotide mutability

5-Methylcytosine (5-Mc) is the most common DNA modification found in eukaryotic genomes. The 5-mC is frequently deaminated to form thymine, which then becomes paired with adenine at DNA replication and therefore becomes a fixed mutation (Ehrlich & Wang, 1981). There is evidence for a specific mismatch repair event which restores a G/T mispair to a G/C pair, protecting the cell from this type of genetic alteration (Brown & Jiricny, 1987), but C to T and G to A transitions, occurring within CpG dinucleotides, are still thought to account for approximately 35% of disease causing mutations (Cooper & Youssoufian, 1988). Three out of nine point mutations detected in this study were found to be such transitions, a rate of 33%, which is compatible with this study.

### 5.4.3 Nucleotide alterations resulting in amino acid substitutions

Four nucleotide changes would be predicted to result in amino acid substitutions in the Btk protein (patients 276, JP, FM and TF), two of which may have occurred as a result of the deamination of a methylated CpG dinucleotide (276 and FM). Two of the patients

(JP and FM) have a severe XLA phenotype and the amino acids which are altered are highly conserved among related tyrosine kinases, indicating the importance of these residues to the functioning of the protein. The other two patients had milder XLA phenotypes (TF and 276) and the amino acids altered are less highly conserved, suggesting that these residues do not have such an essential role in the structure and/or function of Btk. The functional significance of these substitutions will be discussed in sections 7.10 and 7.11.

### 5.4.4 Premature termination codons

Two patients, NAD and AP, had point mutations which resulted in premature stop codons, in exons 8 and 19, respectively. All of the affected boys from both families have the classical XLA phenotype. If the mRNA translated, NAD will have a truncated protein of only 255 amino acids that lacks all of the SH2 domain and kinase domain. There is an increasing body of evidence, however, suggesting that there is a nuclear reduction in the level of mature transcript containing premature termination codons (Urlaub *et al.* 1989) and a concomitant loss of detectable translation in the cytoplasm. This reduction does not seem to occur, however, when the premature stop codon is in the last exon or last third of the penultimate exon of the gene (Cheng *et al.* 1990). Other studies have shown that exons containing premature stop codons are skipped and not included in the mature transcript (Dietz *et al.* 1993). A number of models have been put forward to account for these findings: the "translational translocation" model (Urlaub *et al.* 1989), which involves the translational machinery pulling the transcript into the cytoplasm and being abruptly halted by a premature stop, resulting in degradation of the remaining transcript in the nucleus; the "scanning model", involving nuclear machinery which scans the mRNA for termination codons (Bell *et al.* 1988); and most recently, the "local frame decision" model in which nuclear machinery scans mRNA before splicing, and chooses splice sites on the basis of a long open reading frame in the "exon" (Dietz & Kendzior, 1994). None of these models can accurately account for all the experimental findings, and the actual process, which may involve concepts from all three models, remains to be elucidated. These reports, however, show that premature termination codons must be considered in terms of a reduction in the amount of transcript

and not solely in terms of a truncated protein, unless protein levels have been quantified.

The premature termination in patient AP is in the last exon of the gene and would therefore be expected to produce a truncated protein, lacking the C terminal 24 amino acids including a number of highly conserved residues. There are three severely affected males in this family, suggesting that these last 24 amino acids are essential for the correct expression and/or function of Btk in the developing B cell.

### 5.4.5 Insertions and deletions of one or more nucleotides

Studies of insertional and deletional mutagenesis causing human genetic disease have shown that both processes occur non-randomly and appear to be associated with the local DNA sequence environment (Cooper & Krawczak, 1991; Krawczak & Cooper, 1991). Many insertions and deletions may be explained by direct repeats or runs of single bases promoting slipped mispairing during DNA replication or by inverted repeats or symmetric elements facilitating the formation of secondary-structure intermediates.

Patient LT was found to have an extra A inserted in a run of seven As at nucleotides 341-347. This mutation will result in the protein being out of the correct reading frame from residue 72 and prematurely terminated at residue 83. Runs of the same nucleotide are subject to single base pair insertions and deletions because of slippage of the DNA polymerase during replication (Krawczak & Cooper, 1991). The slippage occurs because of the looping out of a single nucleotide during the replication process - if this nucleotide is on the template strand the result will be the deletion of a single nucleotide, and if it is on the daughter strand, the result will be the insertion of a single nucleotide. This same alteration in this run of seven As has been found in three other XLA patients (Dr H.B. Gaspar, personal communication). The four boys have been shown to come from at least two unrelated families because of the presence of an SSCP polymorphism within the gene (section 6.2.4).

The deletion of a single C in patient FGV, results in the protein being out of frame from residue 207 onwards, with premature termination at residue 216. This premature

termination codon may result in a significant reduction in the amount of mRNA present in the cytoplasm, as described in section 5.4.5. This deletion occurs in a small run of trinucleotide repeats: CCA GC*A GCA GCA CCA, with the asterix indicating the deleted nucleotide. Direct repeats of between two and eight base pairs were found in the region of 98% of the deletions studied by Krawczak *et al.* (1991), although the deletion of a complete repeat element was seen only rarely. The exact mechanism creating these deletions is not known.

Patient AJ has a dinucleotide insertion which may be CA or AC. The exact insertion cannot be identified because of the sequence context of the alteration. The result is that Btk will be translated out of frame from residue 628 onwards, with premature termination at residue 649, in the last exon of the gene. As shown by the premature termination in patient AP, these last residues appear to have an important role in the maintenance of Btk function in the cell. This insertion occurs at TATA*C*CATC and results in TATACACCATC, with the asterix indicating the possible sites of insertion. This AC duplication could be consistent with a slipped misparing model of insertional mutagenesis.

Patient AC was found to have a trinucleotide (AGG) deletion at either residues 1035-1037 or residues 1038-1040. The sequence context of this deletion is GAAAGA(AGG)(AGG)TTTC, with the two possible positions of the deletion in brackets and the flanking complementary repeat underlined. This deletion is compatible with a slipped mispairing model. The consequence of this deletion is the omission of the GGA trinucleotide, coding for glycine 302 in the Btk protein. The functional consequences of this will be discussed in section 7.10.

### 5.4.6 Initiation codon mutations

The initiation codon in eukaryotic mRNA transcripts is usually the first AUG codon. The smaller ribosomal subunit recognises the modified 5' cap of the mature transcript along with several other proteins, and then moves to the AUG codon where the larger ribosomal subunit and other factors associate (Lewin, 1993). A consensus sequence around this region has been derived (Kozak, 1987). The initiator tRNA molecule

carrying the methionine residue is then used to initiate the amino acid chain. Patients SB and BB were found to have nucleotide mutations which result in disruption of the translation initiation codon. Both of these patients have the classical form of XLA, suggesting that disruption of initiation results in a complete lack of functional Btk protein. There are two further potential initiation codons in the gene, at positions 317 and 619, which both have good Kozak consensus sequences (Kozak, 1987), and both in the correct reading frame. Initiation at either of these codons would result in a protein lacking most of the N terminal.

Patient BB is a sporadic case, and was initially diagnosed as having XLA with isolated growth hormone deficiency (Fleischer *et al.* 1980; Sitz *et al.* 1990; Conley *et al.* 1991; Monafo *et al.* 1991). As the underlying defect is possibly a contiguous deletion syndrome encompassing the gene for XLA and another gene involved in growth hormone deficiency, DNA from this patient was initially studied by Southern blot analysis and found to be comparable to normal (results not shown), indicating that he did not have a deletion involving *BTK*. Further studies on this patient showed him to be an XLA patient of short stature (Dr K.Forsyth, Flinders Medical Centre, Adelaide, Australia, personal communication).

### 5.4.7 *BTK* splicing errors

Splicing of heterogeneous nuclear RNA to produce mRNA proceeds by two sequential transesterification reactions, mediated by a spliceosome consisting of five small nuclear RNAs (snRNAs) and numerous proteins (Lamond, 1993). The exact mechanism of catalysis of the reaction has not been defined, but it is known to proceed via the recognition by the snRNAs of the 5' splice site, the 3' splice site and the branch site located 18 to 40 base pairs upstream of the 3' splice site. Cleavage at the 5' splice site is followed by the joining of the 5' end of the intron to the branch site, creating the characteristic lariat. Cleavage then occurs at the 3' splice site and the exon sequences are joined, leaving the intron sequence to be degraded in the cytoplasm. It has been estimated that up to 15% of all point mutations causing genetic disease result in a splicing defect, a large proportion of these involving the invariant GT and AG residues found at

the 5' and 3' ends of introns, respectively (Krawczak *et al.* 1992). Consensus sequences have been derived for the regions around intron/exon boundaries (Pagdett *et al.* 1986) and "Consensus values" which allow the scoring of a potential splice site in relation to the defined consensus, to ascertain the likelihood of use of a cryptic splice site (Shapiro & Senapathy, 1987) have been reported, allowing assessment of the effect of mutations. A splice site mutation can result in an exon no longer being recognised, and therefore excluded from the mature transcript, or the activation of a potential cryptic splice site and the consequent deletion of part of an exon or inclusion of intronic sequences in the mRNA.

The mutation in patient JG was identified initially as an altered *Msp*I restriction fragment, detected by Southern blot analysis of this family (Vetrie *et al.* 1993c). No other alterations were detected on digestion with additional enzymes, suggesting no gross deletion or gene rearrangement. The 21 base pair insertion at the boundary between exons four and five was found, on analysis of recently published intronic sequences, to be the result of an A to G point mutation in the invariant AG splice acceptor sequence. The destruction of this consensus sequence appears to result in the spliceosome selecting a splice acceptor consensus sequence 21 base pairs upstream within intron 4. Using the likelihood scoring of Shapiro and Senepathy ( 1994) the normal splice site at this intron/exon boundary has a score of 90. With the A to G mutation at the -2 position of the 3' splice acceptor sequence, the score is reduced to 74. The cryptic splice site which is utilised in patient JG has a likelihood score of 83, compatible with its preferred usage in the context of the mutation. Splicing would continue as normal with the extra 21 base pairs of intronic sequence remaining in the mRNA and being translated. Although the remainder of this protein is in the correct translational reading frame, the function of the protein appears to be completely disrupted since by this 7 amino acid insertion, as this patient and his two brothers have severe XLA.

Two patients, CAB and CHD, were found to have deletions of the *BTK* cDNA which appear to be the result of a splice site mutation resulting in the deletion of an exon. In patient CAB the deletion of exon 6 results in the substitution of one amino acid and the deletion of 43 amino acids, leaving the reading frame unchanged. In patient CHD the

deletion of exon 18 results in the remainder of the protein being translated out of frame, and prematurely terminated at residue 588. The loss of a complete exon can be attributed to a number of genomic DNA changes involving splice site recognition sequences, and these could now be defined in these patients by analysis of the *BTK* genomic DNA, using the recently published intron sequences (Hageman *et al.* 1994; Ohta *et al.* 1994; Sideras *et al.* 1994). These splice site defects could be due to a point mutation, small deletion or insertion, involving either the splice acceptor sequence or the branch site at the 3' end of the preceding intron, either of which would prevent recognition of the downstream exon. The loss of exon 18 could also be a consequence of a mutation at the 5' splice site sequence of intron 18, a phenomenon which has been taken into account in the formation of the "exon-definition" model of splice site selection (Robberson *et al.* 1990; Niwa *et al.* 1992), which proposes that exons are defined by the identification of the 3' and 5' splice signals at the 5' and 3' ends of the exons, respectively. This model does not, however, take into account the requirement for intron definition to ensure the correct joining of multiple exons or the fact that 5' splice sites closer to the 3' splice site are chosen in preference to those further away (Reed & Maniatis, 1986; Eperon *et al.* 1993). The mechanisms for the selection of splice sites and the high fidelity of the splicing process are complex and not fully understood. The identification of the genomic mutations in these patients which result in this exon skipping, may help in their elucidation.

### 5.4.8 Spectrum of *BTK* mutations detected in this study

Of the fifteen *BTK* mutations detected in this study, four resulted in amino acid substitutions, two resulted in premature termination codons, four involved the insertion or deletion of one or more nucleotides, two nucleotide substitutions resulted in the disruption of the initiation codon and three were the result of splicing errors. Including the six patients shown to have genomic deletions (Lovering *et al.* 1994), these results suggest that the mutations causing XLA are highly heterogeneous. This heterogeneity has implications in the diagnosis of disease and carrier detection, and the understanding of the functions of Btk.

# CHAPTER 6
# MOLECULAR ANALYSIS OF BTK

## 6.1 Introduction

The reasons for identifying disease-causing mutations in a newly isolated gene are threefold; to establish the association between the gene and disease, to provide functional information on the role of the protein in health and disease and to provide more accurate means for disease diagnosis and carrier detection in families affected by the disease. The aim of the work described in this chapter, therefore, was to use the information from the mutation analysis study to further analyse the *BTK* gene and the Btk protein in the context of these three areas.

The *BTK* gene was associated with XLA by Vetrie *et al.* (1993), who identified 8/33 XLA patients with abnormal DNA restriction fragment patterns on Southern blot analysis, using the *BTK* cDNA as a probe. The *BTK* mutation analysis presented in this thesis has shown that mutations in *BTK* are found in most XLA patients studied so far (15/18), confirming the monogenic nature of this disease. The connection between the *BTK* gene and XLA has therefore been firmly established. The mutations causing XLA appear to be highly heterogeneous, suggesting that many different mutations cause the same disease. The possibility remains, however, that Btk could be responsible for other B cell malfunctions, for example less severe abnormalities in B cell growth, differentiation or function, or B cell malignancy. There are patients who have B cell defects less severe than in XLA, and who have apparently normal cell mediated immunity (Dr Alison Jones, personal communication). Other patients have been described who have normal B cell numbers, but abnormal responses to polysaccharide antigens (Umetsu *et al.* 1985; Ambrosino *et al.* 1987), suggesting that they could be the human equivalent of the mouse *xid* defect (see section 1.3.12). Mutations in *BTK* which partially disrupt the protein function could be envisaged to be causative of these milder B cell specific immunodeficiencies.

It has been suggested that Btk may have oncogenic potential on the basis of its similarity to the Src family of protein tyrosine kinases (Vetrie *et al.* 1993c). Also, as Btk is thought to be involved in B cell growth and development, and therefore its dysregulation could be envisaged to result in B cell malignancy. Dysregulation of c-abl tyrosine kinase activity has previously been shown to be directly involved in two types of leukaemia, namely Philadelphia positive common acute lymphocytic leukaemia (ALL) and chronic myeloid leukaemia (CML) (reviewed in Kurzrock *et al.* 1988). ALL is largely a disease of children and has been associated with both genetic and environmental risk factors (Greaves, 1988). Therefore, it could be envisaged that dysregulation of Btk activity could be a cause of some B cell acute leukaemias. With a firm link between mutations in Btk and XLA established therefore, the first aim of this part of this study was to examine the potential links between Btk and other diseases affecting the B cell lineage.

The mutation analysis study presented in Chapter 5 has identified mutations in the *BTK* gene in XLA patients, many of which, on the basis of the homology of Btk to other protein tyrosine kinases, would be predicted to disrupt the function of the protein. Until the protein from the patients can be characterised, however, these functional effects remain only predictions. It may be that, in many cases, the mutation in the *BTK* gene causes either the *BTK* mRNA or the Btk protein to become unstable in the cell, resulting in a complete lack of Btk protein. Studies have shown that lack of Btk protein is found frequently in B cell lines from XLA patients; in the five XLA cell lines studied by Tsukada *et al.* (1993), three had no *BTK* mRNA by northern blot analysis and one had mRNA but no active protein. In other cases, mutations in *BTK* result in the protein having a lack of enzymic activity. One cell line, described by Tsukada *et al.*, which was derived from a patient with a leaky XLA phenotype, had reduced mRNA and reduced active protein. In order to fully characterise the effect of the mutations described in this study on the Btk protein, it was therefore necessary to produce an antibody to Btk. This was generated in this laboratory against a fusion protein containing part of the N terminal of the Btk protein, as detailed in section 2.15.1. This polyclonal antiserum enabled the identification of the Btk protein on western blot analysis of cell lysates and the purification of Btk by immunoprecipitation. Immunoprecipitation of Btk will allow *in vitro* kinase assays to be performed on the protein, to assess Btk enzymic activity. The

correlation between the nucleotide mutation, the protein quantity and activity, and the patient phenotype will be important in the elucidation of the molecular basis of XLA. The second aim of this chapter, therefore, was to begin to study the Btk protein from patients who have defined mutations, to allow correlation between a particular mutation, the encoded Btk protein and the patient phenotype.

Mutation analysis of the *BTK* gene in affected individuals may allow better disease diagnosis, in the case of leaky forms of XLA. Previously, mild B cell immunodeficiencies were not diagnosed as XLA, but this study has shown that some patients with slightly milder immunodeficiency phenotypes have mutations in *BTK* (patients TF and 276, chapter 5), allowing their disease to be more accurately described. *BTK* mutation analysis may also provide the means for more accurate carrier detection in affected families, by assessing the presence of the mutant allele by detecting an SSCP band shift or an altered restriction enzyme digest fragment due to a disease-causing mutation in the *BTK* gene. In families whose disease-causing mutation has not yet been identified, intragenic polymorphisms may provide the means for carrier detection, by allowing the study of the inheritance of the mutant allele. The identification of the *BTK* gene and the development of techniques for quick mutation analysis may improve the accuracy of the information upon which genetic counselling for these families is based. The third aim of this chapter, therefore was to assess these possibilities for XLA carrier detection.

## 6.2 Results

### 6.2.1 Analysis of the *BTK* gene in a patient with a mild B cell immunodeficiency

Patient DW has 5% B cells in his peripheral blood, but virtually no immunoglobulin, as detailed in Appendix II. The *BTK* cDNA from this patient was investigated for mutations by SSCP analysis, as described in section 2.12 and found to be comparable to normal (results not shown).

### 6.2.2 Analysis of the *BTK* gene in two "*xid*-like" patients.

Patients OS and RF had an inability to make antibodies to polysaccharide antigens, as shown by a complete lack of response to vaccinations with pneumococcal polysaccharide (Dr D.Goldblatt, ICH, London, personal communication). Clinical details of these patients are given in Appendix II. In addition, patient OS had two brothers who had died of infection at a young age, while his two sisters are well; a family history which is compatible with an X-linked inheritance. Patient RF also has a brother who died of a chest infection, while his two sisters are well. *BTK* cDNA from both of these boys was analysed by SSCP as described in section 2.12 and found to be comparable to normal (results not shown).

### 6.2.3 Analysis of the *BTK* gene in leukaemia patients

In order to investigate the possibility that *BTK* plays a role in B cell malignancy, the *BTK* cDNA from 23 B cell acute leukaemia patients (ALL), comprising eleven patients with null ALL, five patients with common ALL, four patients with pre-B ALL and three patients with mature B ALL, were studied for mutations by SSCP analysis. No differences were found when they were compared to the normal controls (Katz *et al.* 1994).

### 6.2.4 Detection of a polymorphism in the *BTK* gene

A polymorphism in the *BTK* gene was found which could be detected by SSCP. The band shift was seen in the 119bp fragment generated in reaction A (Figure 6.1a) and sequence analysis of this fragment showed a T to C change at nucleotide 2031 (Figure 6.1b). This nucleotide change is found in codon 633 and changes TGT to TGC. Both of these trinucleotides code for the amino acid cysteine and therefore this polymorphism does not affect the amino acid sequence of the protein. The rarer allele has the published *BTK* sequence, T at position 2031 (Vetrie *et al.* 1993c), and was found in 14 out of 58 XLA and leukaemia patients analysed (this study, Dr H.B.Gaspar and Dr F.Katz, personal communications). The more common allele has C at this position and was

found in the remaining 44 patients. The predicted heterozygosity for this polymorphism is 36%. This nucleotide change does not destroy or create any restriction fragment enzyme site.

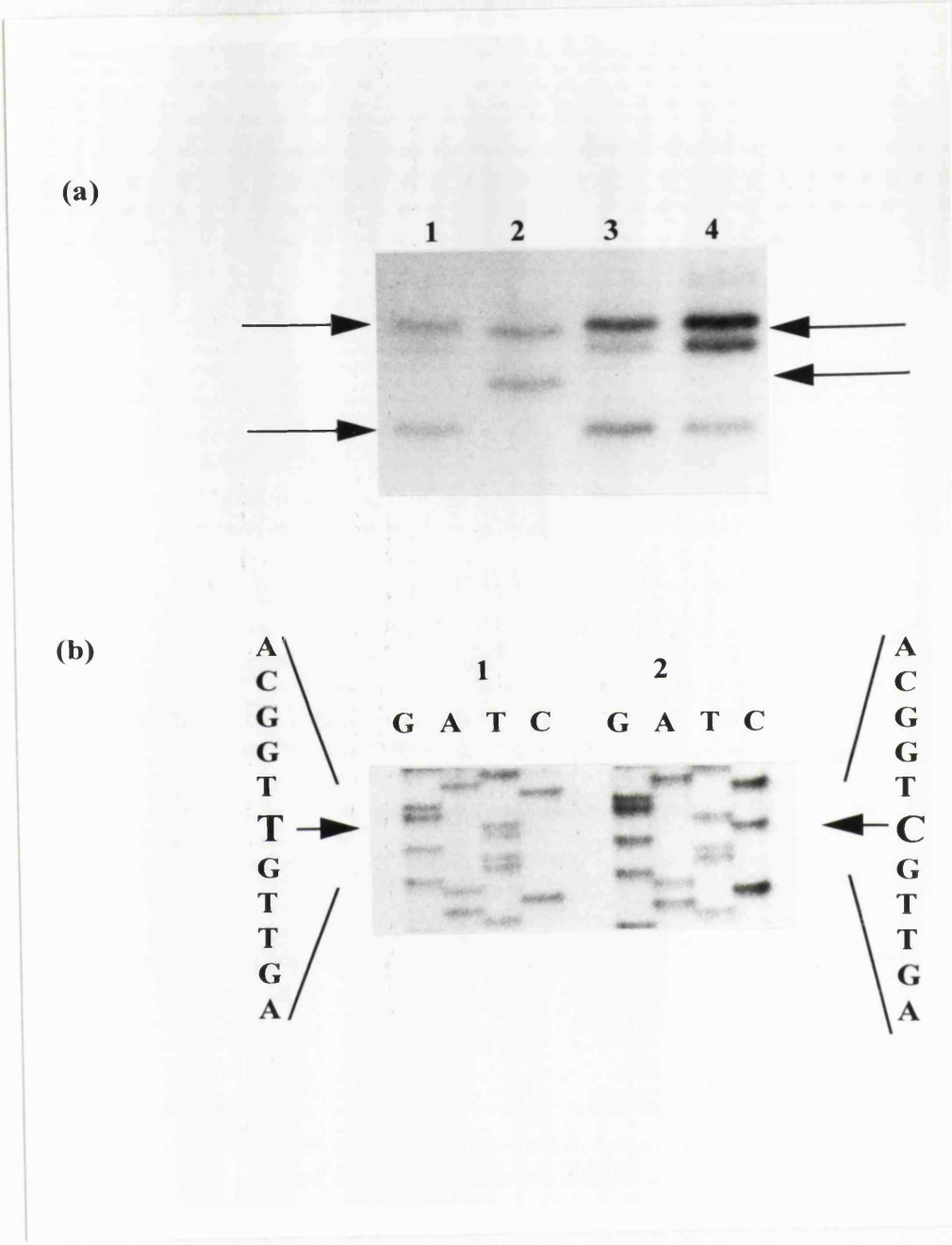
### 6.2.5 The use of direct mutation analysis to improve XLA carrier assessment

The pathological mutation in family P was shown, in this study, to be a G to T transversion at nucleotide 2038 of the *BTK* gene (Chapter 5, patient AP). The mutation was detected by an SSCP band shift in the smaller restriction digest fragment (119bp) produced in reaction A. In order to determine whether the SSCP band shift would be useful in ascertaining carrier status, the same PCR amplification was performed on cDNA from the mother, who was known to be a obligate carrier (Alterman *et al.* 1993), and her two sisters, who were of unknown carrier status. For this analysis, cDNA was generated from both the full mononuclear layer from Ficoll separation of blood and from purified monocytes. Carrier women have a unilateral pattern of X-inactivation in their B cells, but a random pattern in whole blood. The two different cell preparations were therefore used to ascertain any differences in the ability to detect the mutant allele. The samples were analysed on an SSCP gel. No differences could be seen when the samples from the two cell preparations were compared The SSCP band shift could be seen when the affected boy and the normal control were compared (Figure 6.2b). The obligate carrier mother could be seen to have both the normal and mutant bands, but the bands from the two sisters could not be resolved (Figure 6.2b). As this nucleotide change results in the destruction of a *Nla*III site, the PCR products were digested and analysed on a polyacrylamide gel as described in section 2.6.3, and the result shown in Figure 6.2c. A schematic diagram of the digest is shown in Figure 6.2d. The loss of the restriction site gave an larger 75bp fragment in the affected boy which was not seen in the normal controls, which only had the smaller 59bp fragment. The obligate carrier and her two sisters all had both bands indicating that all three are carriers of the mutant allele.

### 6.2.6 The isolation of Btk protein from XLA patients

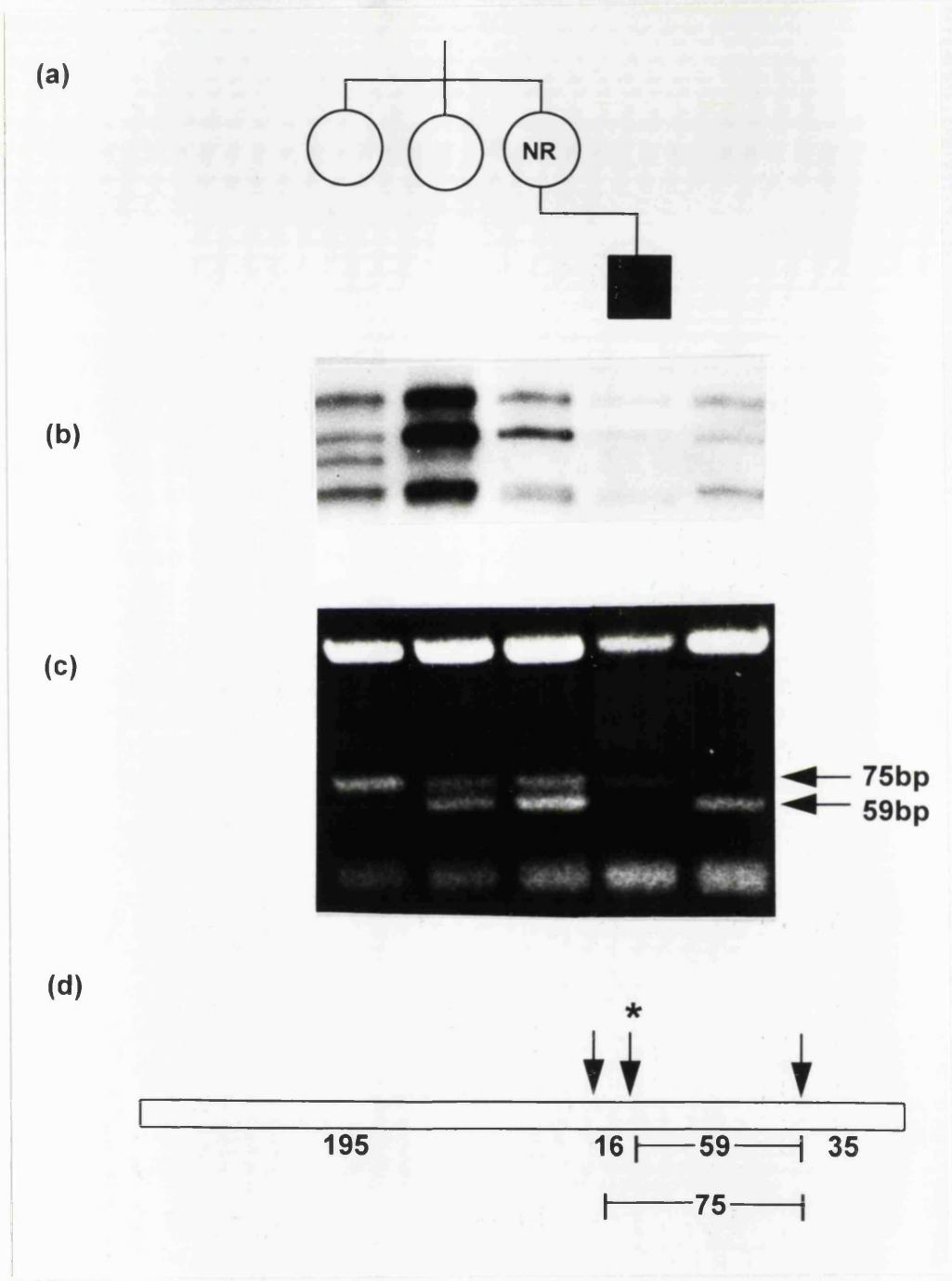
Analysis of the Btk protein from XLA patients was found to be hampered by the lack of

**Figure 6.1 Identification of a polymorphism in the *BTK* gene**




(a) SSCP analysiis of the 119bp fragment from reaction A of the *BTK* gene (as detailed in section 2.12) from four unrelated individuals. Tracks 1, 3 and 4 show the more common SSCP pattern and track 2 shows the rarer pattern (see text)) and (b) sequence analysis of (1) an individual with the rarer SSCP pattern showing a T at position 2031 of the *BTK* gene and (2) an individual with the more common SSCP pattern showing a C at position 2031 of the *BTK* gene

**Figure 6.2 Determination of carrier status by direct mutation analysis**




Carrier determination in family P showing; (a) pedigree of family, (b) SSCP analysis of cDNA from each individual showing the band shift in the affected boy but ambiguous carrier detection. The upper and lower bands are the single stranded DNA and the band in the middle is the double stranded DNA. The track on the left also shows and additional band which cannot be identified. (c) restriction digest pattern showing the normal allele as a 59bp fragment and the mutant allele as the 75bp fragment and (d) a schematic representation of the *Nla*III restriction sites (indicated by arrows) around the site of the mutation, with the site lost in the mutant allele indicated by a star.

availability of cells containing the protein. EBV transformed B cell lines provide the best source of material, but as XLA boys have virtually no B cells in their peripheral blood, the establishment of these cell lines is difficult. Monocytes purified from the mononuclear cell layer of Ficoll separated blood can be used as a source of Btk protein but 50 ml of blood must be analysed on the same day as the sample is taken, for sufficient monocytes to be isolated; only half the number of monocytes could be isolated from blood left at room temperature overnight (results not shown). Repeated attempts to study Btk from Ficoll purified mononuclear cells stored in liquid nitrogen proved unsuccessful. This restricted analysis of the Btk protein of XLA patients to those who were present in the Hospital, or those from whom an EBV cell line could be generated.

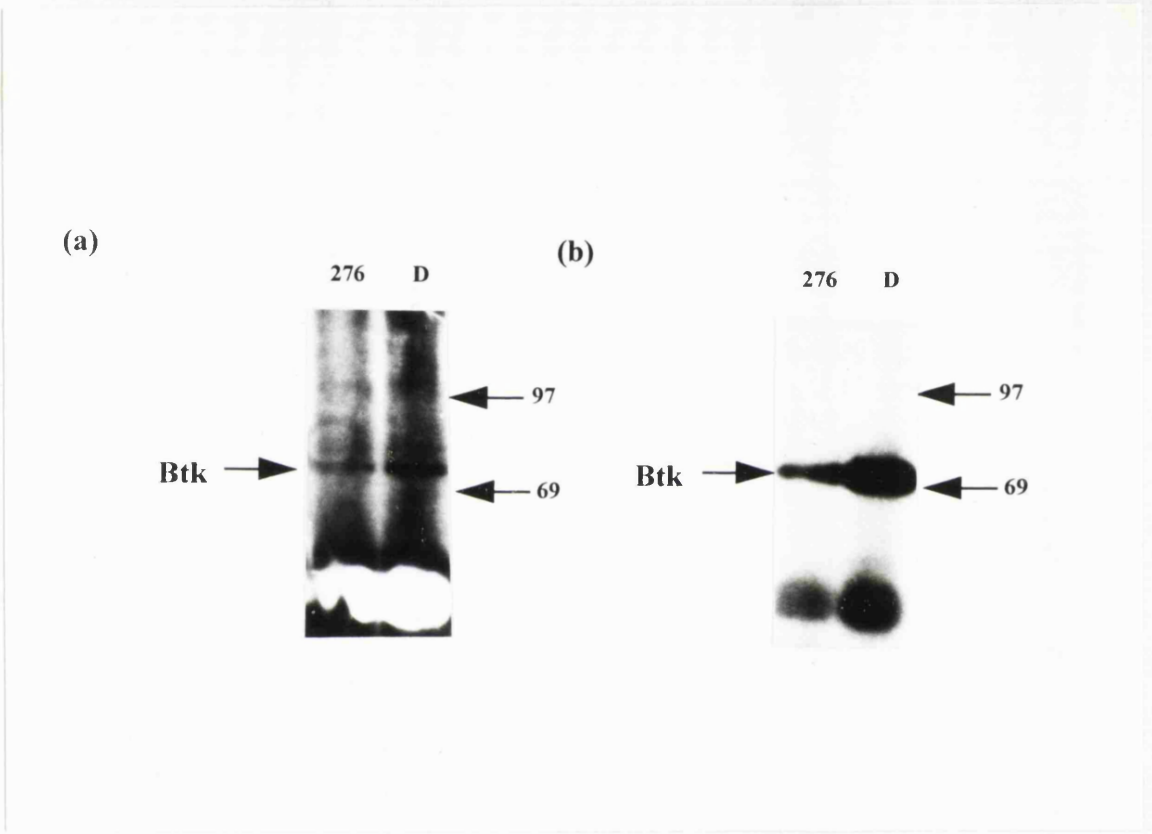
### 6.2.6.1 Analysis of levels of protein and kinase activity in patient 276

A B cell line derived from EBV transformation of bone marrow from patient 276 (a gift of Dr M.de Weers, Leiden, The Netherlands) was analysed for the presence of Btk protein. This patient was shown to have a mutation in the SH2 domain, arginine 288 changed to tryptophan, of the Btk protein (section 5.3.4). Western blot analysis using the Btk antiserum on a hypotonic lysate of these transformed B cells showed that Btk was present (Figure 6.3a). Immunoprecipitation of Btk from these cells, followed by a kinase assay (described in section 2.15) showed that the protein had enzyme activity comparable to a Daudi B cell line (Figure 6.3b).

### 6.2.6.2 Analysis of Btk protein in patients CP and CHD

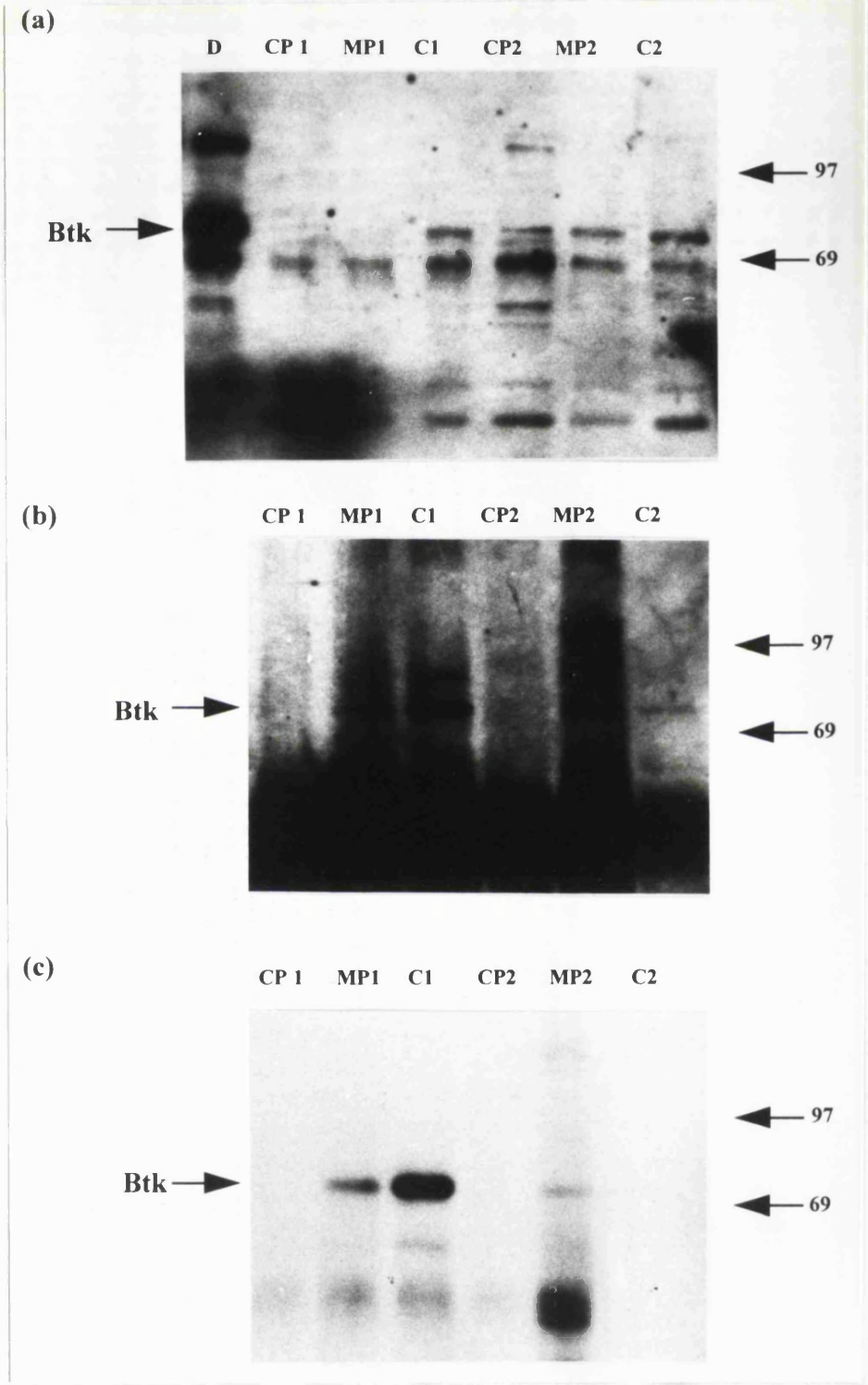
The XLA phenotype in family P had been found to be caused by a premature stop codon causing the protein to be terminated 24 amino acids from the end. This protein would be predicted to be approximately 70kD compared to the 77kD normal Btk. Monocytes were purified from the blood of patient CP, his obligate carrier mother and a normal control. The monocytes and the remaining mononuclear cells (predominantly B cells and T cells) were subjected to hypotonic lysis. Western blot analysis of these samples using the Btk antiserum showed the presence of the 77kD Btk protein in the monocytes and T/B cells of both the mother (MP) and the normal control and a band of the predicted size

**Figure 6.3 Analysis of Btk protein in patient 276**




(a) Western blot analysis of the Btk protein from an EBV transformed B cell line from patient 276 and a Daudi cell control, with the 77kD Btk protein indicated by an arrow on the left and (b) an autoradiograph of an *in vitro* kinase assay showing the activity of the Btk protein from patient 276 compared to a Daudi control. Sizes of markers are given in kD and are indicated by arrows on the right.

**Figure 6.4 Analysis of Btk protein from patient CP**




(a) Western blot analysis of hypotonic lysates, using the Btk antisera. Samples shown are Daudi cell line (D), non-adherent mononuclear cells from the peripheral blood of XLA patient CP (CP1), his obligate carrier mother MP (MP1) and a normal control C (C1) and monocytes from the peripheral blood of patient CP (CP2), his mother (MP2) and normal control C (C2). (b) Anti-Btk immunoprecipitates from the same cell types CP1, MP1, C1, CP2, MP2 and C2 . (c) Autorad showing the kinase activity of the anti-Btk immunoprecipitates shown in (b). Size markers are shown on the right in kD.

(70kD) in the monocytes of patient CP (Figure 6.4a). Additional cross-contaminating bands, which were also seen in the Daudi control, were seen in this track.

The Btk protein was immunoprecipitated from additional samples of the same types of cells from these three individuals, and subjected to a kinase assay as described in section 2.15.5. Western blot analysis of these samples showed the presence of the 77kD Btk protein in both samples from the mother (MP) and the normal control, but the absence of this protein in both samples from the affected boy (Figure 6.4b). Autoradiographic analysis of the membrane showed that the 77kD protein in the samples from the mother and the normal control had detectable *in vitro* kinase activity, but this activity was absent from the samples from the affected boy (Figure 6.4c). This analysis was performed twice, on separate occasions, with the same results.

The same analysis was performed on blood from patient CHD. The XLA phenotype in this patient was shown in this study to be caused by the deletion of exon 18 (Chapter 5). The results from this analysis of the Btk protein showed the complete absence of detectable protein as shown by western blot analysis of lysates, immunoprecipitation and kinase assays (results not shown). It became apparent, however, that the results of the immunoprecipitation of Btk from the monocytes of normal controls was not consistent.

## 6.3 Discussion

### 6.3.1 Involvement of *BTK* in other B cell immunodeficiencies

This study has involved a preliminary investigation into the potential involvement of *BTK* in disease phenotypes other than XLA. It is possible that mutations in *BTK* may be involved in mild forms of B cell immunodeficiencies, which may not always be clinically diagnosed as XLA. Amino acid substitutions at the less conserved residues in the protein may only have mild effects on the protein function. The case of hypogammaglobulinemia studied here did not have any alterations in the *BTK* gene by SSCP analysis, but another patient who was initially reported as having 5% B cells has been found to have a SSCP

band shift in the 195bp fragment of reaction A (Dr H.B.Gaspar, personal communication) indicating the presence of a mutation. The exact nucleotide change has not yet been elucidated. This fragment is the same as the one containing a mutation in Family F, who showed a milder form of XLA (Section 5.3.1).

Two patients with *xid*-like phenotypes were investigated for mutations in *BTK*. No mutations were found in either of these patients by SSCP analysis. This does not completely exclude the possibility of the involvement of *BTK* in this phenotype as SSCP does not detect 100% of mutations (see section 5.2.1). Recently, however, a severely affected XLA patient has been described with amino acid substitutions at the same residue as the substitution found in the *xid* mouse (de Weers *et al.* 1994b). Amino acid 28 is changed to histidine in the XLA patient and cysteine in the *xid* mouse. Histidine and cysteine are thought of as very similar amino acids, and therefore it could be predicted that these amino acid substitutions will not have vastly different phenotypic effects. The differences in immunological phenotype may therefore be due to the different mechanisms of B cell ontogeny in the two species.

A patient previously diagnosed as having common variable immunodeficiency was found to have a genomic deletion of the *BTK* gene by Southern blot analysis (Vorechovsky *et al.* 1993), illustrating that patients diagnosed as having immunodeficiencies other than XLA may have mutations in this gene.

### 6.3.2 Btk and B cell leukaemias

The potential involvement of Btk in some types of B cell leukaemia is still to be elucidated. No evidence for *BTK* mutations in twenty-three B cell leukaemias was found in this study by SSCP analysis, or any evidence for changes in the level of Btk protein in twelve of these patients (Katz *et al.* 1994). There are many documented mutations in c-Src which have been reported to increase activity and produce a transforming phenotype. These are found in several regions of the gene including the C terminal regulatory tyrosine (Y527), the SH2 domain and the SH3 domain (Hirai & Varmus, 1990; O'Brien *et al.* 1990; Superti Furga *et al.* 1993). Deletions of the SH2 or SH3 also

seem to activate the transforming potential of the protein, as does alterations to the spacing between the C terminal regulatory tyrosine and the rest of the kinase domain. Other mutations in these regions appear to reduce the transforming activity of activated c-Src, suggesting that there are both positive and negative regulatory mechanisms associated with the SH2 and SH3 domains (Hirai & Varmus, 1990; O'Brien *et al.* 1990). The proposed mechanism for the regulation of c-Src activity involves the SH2 domain binding the phosphorylated C terminal tyrosine, bringing the kinase domain into an inactive conformation. Dephosphorylation of this residue, mutation of the tyrosine to phenylalanine or mutation of certain residues in the SH2 domain appear to release this control mechanism (Liu *et al.* 1993). Btk is related to c-Src, although it lacks the C terminal tyrosine and therefore may be subject to a different mechanism of control. Deletion of Btk SH3 domain or mutation of SH2 residues critical in phosphotyrosine binding does not seem to have activating effects on the Btk protein, but result in the disruption of the function of Btk and lead to XLA. It is possible that other types of B cell malignancy may involve Btk activation, for example chronic lymphocytic leukaemia or follicular lymphomas, and therefore a larger study characterising Btk expression in these malignancies may prove to be interesting. Although there is no evidence at present, therefore, to link Btk with B cell leukaemia, until the mechanisms of Btk activity have been fully elucidated, the possibility of a connection still remains.

### 6.3.3 Analysis of Btk protein

This study has shown that the Btk protein from patient 276 has kinase activity, as shown by an *in vitro* kinase assay. The XLA phenotype in this patient is caused by a single amino acid substitution in the SH2 domain, arginine 255 being changed to tryptophan. This suggests that this SH2 domain mutation does not have a detectable effect on the enzymic activity of the protein. The SH2 domain mutant reported by Saffran *et al.* (1994) was found to be synthesised at the normal rate as shown by metabolic labelling experiments, but was found to be quickly degraded suggesting that Btk instability was one of the main consequences of the mutation. This emphasises that predicted effects of mutation must be followed by studies on the stability and activity of the protein.

Btk protein of the 70kD truncated size was seen on western blot analysis of a hypotonic cell lysate of monocytes from patient CP, but could not be detected by western blot analysis of anti-Btk immunoprecipitates. No Btk protein could be detected in monocytes from patients CHD, either by western blot analysis of a hypotonic cell lysate or by immunoprecipitation of Btk. The immunoprecipitation results from normal controls appeared to be inconsistent, suggesting that purified monocytes may not be suitable for Btk protein analysis. Btk is thought to be expressed at similar levels in B cells and in myeloid cells (de Weers *et al.* 1993; Smith *et al.* 1994a; Genevier *et al.* 1994a). This protein is not essential for myeloid cell growth and/or function, as shown by the lack of identifiable myeloid cell dysfunction in XLA patients and the role of Btk in these cells is unclear.

No Btk protein was found in a cell line produced from another XLA patient (Genevier *et al.* 1994a). This patient had been found to have a point mutation resulting in the substitution of threonine 33 by proline. It is possible that many amino acid alterations result in unstable proteins which are degraded quickly. Analysis of the presence of Btk protein in further XLA patients may allow confirmation of the predicted effect of the mutation on the protein, and analysis of some of the inter-relationships between different domains of the protein.

### 6.3.4 Implications of *BTK* mutation analysis for carrier detection in XLA

Direct mutation analysis may result in much improved carrier detection for XLA by permitting the direct detection of the mutant allele. In the family studied here, the SSCP band shift was not sufficiently large to clearly ascertain carrier status. However, in patients where the band shift indicating the mutation is larger, it should be possible to unequivocally determine carrier status using SSCP analysis. Carrier determination by SSCP analysis will be made easier when genomic DNA instead of cDNA can be used as the starting material, as both alleles will be present at exactly the same level, compared to cDNA where levels are subject to variations due to of X-inactivation patterns. In the family studied here, however, the mutation causing the disease results in the disruption of a *Nla*III site allowing the identification of all three women as carriers of XLA.

Several of the XLA patients analysed in this study, have nucleotide changes which alter restriction enzyme digest sites (Table 5.1) which can potentially be used for simple and quick detection of the mutant allele. These changed restriction patterns will also be useful for pre-natal diagnosis when the gene can be amplified from genomic DNA and subjected to digestion by the appropriate restriction enzyme.

SSCP analysis of the *BTK* gene in this study revealed an intragenic SSCP polymorphism which has an predicted heterozygosity of 36%. This polymorphic C or T at position 2031 of the *BTK* gene does not result in the creation or destruction of any restriction enzyme site, preventing its detection by analysis of restriction digest patterns. The detection of this polymorphism will therefore have to be by SSCP analysis or DNA sequencing. Using purified mononuclear cells, SSCP analysis may show the presence of the two alleles in heterozygous women. The allele co-inherited with the disease must be the mutant allele, as this polymorphism is intragenic. Analysis of this polymorphism has proved to be useful in the analysis of four XLA patients with the same mutation (section 5.4.6), as two of the patients had one allele and two had the other allele, suggesting that these patients were from at least two separate families. Another polymorphism detectable by SSCP analysis has been reported in the 3' untranslated region of the gene (Vorechovsky, 1994).

# CHAPTER 7
# DISCUSSION

## 7.1 The molecular basis of XLA

An understanding of the molecular basis of XLA requires full knowledge of the gene and the protein involved in the disease phenotype. This requires the isolation of the gene, knowledge of its DNA sequence and its mechanisms of control of expression, a study of the mutations which result in the disease, and information on the structure and function of the encoded protein. At the start of this study, the gene responsible for XLA had not been identified, but by using a positional cloning approach, the location of the gene was narrowed down to an approximately 2-4cM region in Xq22 (Lovering *et al.* 1993a). Since then, the *BTK* gene has been identified and shown to be responsible for the disease, and the molecular basis of XLA is beginning to be understood.

## 7.2 Mapping in Xq22

The identification of the gene responsible for XLA marked the end of a long search by several laboratories. The *BTK* gene lies 200kb proximal to DXS178, the locus mapping the closest to XLA by genetic linkage analysis, and has been placed in the following order of loci in Xq22: cen-DXS442-*BTK*-*GLA*-DXS178-DXS101-tel (Vetrie *et al.* 1993c). *BTK* has been physically mapped 50-70kb proximal to *GLA*. The physical distance between *BTK* and DXS442 is not known, but must be greater than 800kb (Sweatman *et al.* 1994). During the studies which lead to the isolation of this gene, physical maps (Vetrie *et al.* 1993b; Sweatman *et al.* 1994) and genetic maps (Lovering *et al.* 1993a; Parolini *et al.* 1993) of Xq22 have been constructed which may facilitate the identification of other genes in the region. The region has now been represented in several YAC contigs (this study; Vetrie *et al.* 1994; Parolini *et al.* 1993). Ten or more CpG islands, indicative of transcribed sequences, have been described in the region between DXS442 and DXS101 (Vetrie *et al.* 1993a; Vetrie *et al.* 1993b; O'Reilly *et al.* 1993a; Sweatman

*et al.* 1994) and four other cDNAs were isolated in conjunction with *BTK* (Vetrie *et al.* 1993c; Vorechovsky *et al.* 1994). This evidence suggests the presence of several other genes in the region. The genes responsible for several genetic diseases have been mapped to this region, including X-linked megalocornea (Bleeker-Wagemakers *et al.* 1991) and FG syndrome (Zhu *et al.* 1991). There are no cytogenetically detectable deletions described in XLA patients and all the deletions studied so far only extend beyond the 3' region of the gene. The $\alpha$ galactosidase gene, responsible for Fabry disease, has a very low rate of deletion (Bernstein *et al.* 1989) and, despite its close proximity to *BTK* (within 90kb), no patient has been reported suffering from Fabry disease and XLA. This could suggest that deletion of this region is not compatible with survival, indicative of the presence of essential gene(s), probably distal to *BTK* (at the 5' end of the gene). The data produced during the search for the gene responsible for XLA will aid in the identification of these other genes in the Xq22 region.

## 7.3 The involvement of *BTK* in immunodeficiency

### 7.3.1 XLA

The identification of the *BTK* gene has resulted in the ability to study XLA patients for mutations and to begin to understand the molecular basis of the disease in each patient. Most XLA patients have detectable mutations in the coding region of the gene, and further analysis of the remainder (in this study, 3/18) may define mutations in the promoter or other control sequences. Thus, mutations in *BTK* have been definitively associated with XLA. In addition, a number of patients with milder phenotypes have been found to have mutations in *BTK* (this study; Saffran *et al.* 1994; Ohta *et al.* 1994), illustrating that other patients with B cell restricted immunodeficiencies, which have not previously been diagnosed as XLA, may have a mild form of this disease. The molecular basis of the XLA-like disease seen in two females (Conley and Sweinberg, 1992) remains to be elucidated.

### 7.3.2 XLA and isolated growth hormone deficiency

The identification of the *BTK* gene has allowed further study of families affected by XLA and isolated growth hormone deficiency (IGHD). Vorechovsky *et al.* (1994) examined DNA from three families affected by XLA/IGHD by Southern blot analysis, using the *BTK* cDNA as a probe, in order to identify any genomic deletions which may involve an adjacent gene. No evidence was found to suggest any such deletions, therefore excluding the possibility of contiguous deletions as the basis for the coinheritance of these disorders. Conley *et al.* (1994) have reported finding mutations in the *BTK* gene in affected individuals from two families previously documented to have XLA/IGHD (Conley *et al.* 1991; Sitz *et al.* 1990). Linkage analysis of these families had shown that the disease locus mapped to the same region as XLA. Mutational analysis has now shown that the disease causing mutation in one family is a premature stop codon at residue 375 in the SH2 domain. The mutation in the second family was found to result in a leucine to proline substitution at position 542 in the kinase domain. These mutations do not have any unique characteristics which could be associated with IGHD, suggesting that the XLA and the IGHD are not genetically related. A recent study has suggested that true cases of XLA/IGHD are rarer than currently assumed as many of the reported cases of XLA/IGHD can be accounted for by growth hormone insufficiency due to delayed growth and puberty (Buzi *et al.* 1994). Testosterone priming of subjects can produce normal growth hormone levels. The molecular basis for the association of XLA and IGHD remains unknown.

## 7.4 The *BTK* gene

The 2.5kb *BTK* mRNA contains a 1.9kb open reading frame, beginning with an initiation codon at nucleotide 133. The genomic structure of the *BTK* gene has recently been established by several groups (Ohta *et al.* 1994; Hageman *et al.* 1994; Sideras *et al.* 1994). The gene is encoded in 19 exons and is contained within 37.5kb of genomic DNA, with the first exon containing only 5' untranslated region (UTR). The coding exons in the human gene vary in size from 55bp to 503bp and the intron sizes vary from

179bp to 9.2kb. The promoter region of the *BTK* gene has been analysed and no obvious TATAA or CAAT-like sequences upstream of the transcription start site (Hageman *et al.* 1994; Sideras *et al.* 1994), indicating that the transcription factors which bind to these motifs do not play a role in *BTK* expression. The sequence around the start site in *BTK* is identical in seven out of eight nucleotides to the "initiator" sequence (Smale & Baltimore, 1989). This is a promoter element which acts, in concert with or independently of the TATA box, to initiate transcription. The intron/exon boundaries of the *BTK* gene have been found to be in similar positions in the mouse gene, and the gene has been found to cover approximately 43.5kb, making it slightly larger than the human gene (Sideras *et al.* 1994).

## 7.5 Expression of *BTK*

The expression of *BTK* has been studied by northern blot analysis and by western blot analysis using an anti-Btk antisera. *BTK* expression appears to be limited to the haematopoietic lineages. Expression has been detected in B lineage cells, from pro-B cell to mature B cells, but not in plasma cells (de Weers *et al.* 1993; Smith *et al.* 1994a; Genevier *et al.* 1994). It has also been detected in myeloid lineage, erythroid lineage and mast cells (de Weers *et al.* 1993; Smith *et al.* 1994a; Genevier *et al.* 1994). No expression was detected in T cell lines or normal adult tonsillar T cells. Btk has also been detected at normal or reduced levels in cells derived from XLA patients (this study; Tsukada *et al.* 1993; Genevier *et al.* 1994).

## 7.6 The Btk protein

The *BTK* mRNA is translated into a 659 amino acid protein of 77kD. Database searches showed that both the gene and the protein have significant levels of homology to the Src family of protein tyrosine kinases (PTKs), suggesting that Btk is a member of this family (Vetrie *et al.* 1993c). Particularly high regions of homology between Btk and Src encode

the characterised Src homology (SH) domains, SH1, SH2 and SH3 (Figure 7.1). Within the Src family, Btk appears to be the most closely related to the proteins Itk (Siliciano *et al.* 1992) and Tec (Mano *et al.* 1993) and it has been postulated that these three proteins make up a subfamily of PTKs. These three proteins show high levels of homology to Src over some regions of the protein, but exhibit a number of differences. They lack the SH4 domain found at the N terminal of Src which contains the myristoylation site, thought to play a role in localising Src to the plasma membrane, and have a much longer N terminal domain, part of which shows a high level of homology to the recently described Pleckstrin homology (PH) domain (Musacchio *et al.* 1993). This long N terminal domain also contains a proline rich region, recently named the Tec homology domain, because of the level of conservation between Tec, Btk and Itk (Smith *et al.* 1994b). The Btk subfamily of PTKs lack the C terminal tyrosine residue, thought to be involved in the regulation of Src enzymic activity (see section 6.3.2). On the basis of the homology of Btk to other proteins and the definition of previously characterised domains, Btk can be considered to be composed of a number of modular domains as shown in Figure 7.1. This allows the assignment of certain functions to particular regions of the protein and the speculation of the effect of mutations with the regions of the gene encoding these domains.

## 7.7 The pleckstrin homology domain

The mutation in patient JG (section 5.3.6) would result in a seven amino acid insertion in the PH domain of the protein. This domain is a region of approximately 100 amino acids, first detected as an internal repeat in pleckstrin, a 47 kDa protein which is the major substrate of protein kinase C in platelets (Haslam *et al.* 1993; Mayer *et al.* 1993). This domain has now been found in proteins with diverse cellular functions (Musacchio *et al.* 1993) many of which are thought to be involved in cell signalling. Some cytoskeletal proteins including dynamin and $\beta$-spectrin have also been found to contain PH domains. Several of these proteins also have SH2 and SH3 domains. The PH domain is thought to be too small to have a catalytic function of its own and may be involved in molecular recognition (Gibson *et al.* 1994). It is more divergent than SH2

**Figure 1.3 Diagrammatic representation of Btk**


The 659 aa Btk protein has SH1, SH2 and SH3 domains and a long N terminal domain, thought to contain a pleckstrin homolgy (PH) domain and a Tec homolgy (TH) region

and SH3 domains and, like SH2 and SH3 domains, appears to be modular, functioning independently of its position in the protein. Three dimensional nuclear magnetic resonance spectroscopy of the pleckstrin PH domain has shown that the structure of this domain is a $\beta$ barrel of seven anti-parallel $\beta$ strands (a-g) with a C terminal amphiphilic $\alpha$ helix cap (Yoon *et al.* 1994). A very similar structure has been found for the $\beta$ spectrin PH domain, illustrated in Figure 7.2, with the differences being found in the loop region of the protein (Macias *et al.* 1994). There appears to be a cluster of positive charges around with what may be a binding cleft, which is associated with strands a-b and e-f. PH domains have recently been shown to interact with a number of other proteins *in vitro*. The N terminal part of the PH domain from pleckstrin has been shown to bind to phosphatidylinositol-4,5-biphosphate (PIP2) (Harlan *et al.* 1994). Studies of the $\beta$ adrenergic receptor kinase ($\beta$ARK) PH domain suggested that the region could bind the $\beta\gamma$ subunits of trimeric G proteins (Touhara *et al.* 1994), and a recent study by Yao *et al.* (1994) showed that the Btk PH domain, and the PH domains of the other proteins of the Btk subfamily, bind to protein kinase C (PKC) in mast cells.

The twenty one base pair insertion in the *BTK* gene of an XLA patient described in this study (section 5.3.6) would result in a seven amino acid insertion at position 66 of the PH domain (nomenclature as in Gibson *et al.* 1994), in the middle of strand f (see Figure 7.2). This strand is thought to be involved in the putative binding cleft and therefore an insertion such as this could be predicted to disrupt binding activity of the PH domain. Other recently described mutations in the Btk PH domain include amino acid substitutions at position 28 of Btk, changing arginine to cysteine in the *xid* mouse (Thomas *et al.* 1993; Rawlings *et al.* 1993) and from arginine to histidine in a severely affected XLA patient (de Weers *et al.* 1994b). This residue is found in strand b (PH domain position 20) and is thought to be at the bottom of the positive binding pocket, and likely to be involved in ligand binding (see Figure 7.2). Northern blot analysis of *BTK* mRNA, metabolic labelling of the Btk protein and autokinase assays of immunoprecipitated Btk protein showed no differences between *xid* B cells and normal mouse B cells indicating that the *xid* mutation did not affect expression of the gene, stability of the protein or the autokinase activity of the protein (Thomas *et al.* 1993; Rawlings *et al.* 1993). The Btk PH domain binding of PKC in mast cells was shown to be reduced by the presence of the

*xid* amino acid substitution (Yao *et al.* 1994). Another mutation resulting in a threonine to proline substitution has been reported at Btk residue 33, in the PH domain. This mutation has been shown to result in the complete absence of Btk protein in an EBV cell line derived from this patient, suggesting that it makes the protein unstable (Genevier *et al.* 1994).

In conclusion, the Btk PH domain appears to be a modular domain involved in protein - protein interactions with one or more substrates, including PKC. Disruption of the PH domain by the insertion of seven amino acids in strand f or the substitution of residue 28 by cysteine or histidine appears to completely disrupt the functioning of Btk, resulting in the human XLA or mouse *xid* phenotype. PH domain interactions must therefore be critical to the normal functioning of this protein.

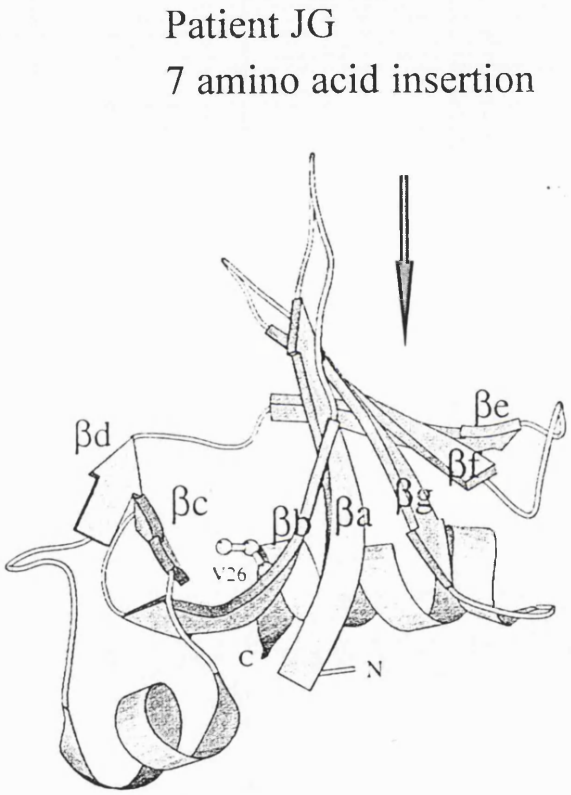
## 7.8 The Tec homology domain

The region between the PH domain and the SH3 domain is of unknown function but has tentatively been designated the Tec homology (TH) domain (Smith *et al.* 1994b). There are two proline-rich regions in this domain which resemble the SH3 binding site consensus sequence (Ren *et al.* 1993; Yu *et al.* 1994), residues 185-192 KKPLPPTP and residues 199-206 KKPLPPEP, suggesting that this domain may mediate interactions with the SH3 domain(s) of other molecules. A recent study has shown that the SH3 domains of the Src family PTKs Fyn, Lyn and Hck bind to these proline rich sequences in Btk (Cheng *et al.* 1994). No XLA patients have been found with amino acid substitutions in this region of the *BTK* gene, in this or any other study. The significance of this, if any, is not yet known.

## 7.9 SH3 domains

SH3 domains are modular regions of approximately sixty amino acids which are thought to be involved in protein-protein interactions resulting in the mediation of the subcellular

**Figure 7.2 Backbone ribbon diagram representation of a PH domain**




Representation of the $\beta$-spectrin PH domain structure, taken from Gibson *et al.* (1994). The $\beta$ strands are labelled $\beta$a-$\beta$g. The carboxy C terminus is at the lower back and the amino terminus is at the bottom centre. V26 is the residue corresponding to the site of the amino acid substitution in the *xid* mouse (Thomas *et al.* 1993; Rawlings *et al.* 1993) and in a severely affected XLA patient (de Weers *et al.* 1994b).

location of proteins and the transduction of intracellular signals (reviewed in Mayer & Baltimore, 1994). This domain is found in many proteins involved in cell signalling, for example, the Src family of tyrosine kinases and phospholipase C$\gamma$, and in cytoskeletal proteins, for example, dynamin. SH3 domains are often found in conjunction with SH2 domains although their functions are entirely separate. Three dimensional NMR analysis of the Src SH3 domain suggests that the secondary structure is formed of two three-stranded anti-parallel $\beta$ sheets, with a hydrophobic core at the interface (Yu *et al.* 1992). This provides the binding site for the consensus recognition sequence of SH3 binding proteins. The recognition sequence has been shown by mutational studies to be a nine or ten amino acid stretch which is rich in proline residues (Ren *et al.* 1993). The binding site of the SH3 domain contains some highly conserved residues but the lack of total conservation suggests that there may be different binding specificities for each SH3 domain. This has now been demonstrated by *in vitro* binding studies (Gout *et al.* 1993). It has recently been shown that the Btk SH3 domain does not have affinity for its own SH3 binding motifs, present in the Btk Tec homology region (Cheng *et al.* 1994). Experiments with the GTPase dynamin showed that binding to SH3 domain activated the GTPase activity, suggesting that SH3 domain binding may exert some regulatory effects (Gout *et al.* 1993).

No amino acid substitutions have been found in the Btk SH3 domain, in any study so far. This may simply because they have not yet been identified, but it also may be because amino acid substitutions in this domain do not severely disrupt the functioning of the Btk protein.

## 7.10 SH2 domains

Three patients were found to have *BTK* mutations in the part of the gene encoding the SH2 domain. SH2 domains have been shown to bind specifically and with high affinity to phosphorylated tyrosine residues, and are therefore thought to link proteins into signal transduction networks involving tyrosine phosphorylation (reviewed in Mayer & Baltimore, 1994). They have been found in a wide variety of proteins including non-

receptor tyrosine kinases, phospholipase C$\gamma$, Ras GTPase-activating protein (GAP) and adaptor proteins, sometimes in duplicate and sometimes in conjunction with an SH3 domain. X-ray crystallographic studies of Src SH2 domain have shown that its secondary structure consists of two antiparallel $\beta$ sheets and two surrounding $\alpha$ helices, which together form a binding pocket into which the phosphorylated tyrosine is inserted (Waksman *et al.* 1992, see Figure 7.3a). For ease of comparison of different SH2 domains, a consensus nomenclature has been proposed (Eck *et al.* 1993). Each residue is numbered within its individual secondary structure, with the $\beta$ strands named $\beta$A - $\beta$G and the $\alpha$ helices named $\alpha$A and $\alpha$B. The loops between the strands and the helices are named using the names of the secondary structures which they join. Three conserved amino acids found in this pocket interact with the phosphotyrosine, as shown in Figure 7.3b. A second pocket formed by the SH2 domain is thought to be involved in the specificity of the interaction between the SH2 domain and the phosphopeptide. This pocket binds the residue found at the +3 position after the phosphorylated tyrosine (Waksman *et al.* 1992; Waksman *et al.* 1993). SH2 domains from different proteins are thought to recognise phosphotyrosine containing proteins with different specificities; for example, the SH2 domains of the Src family PTKs are thought to recognise phosphotyrosine in the sequence context of leucine or isoleucine at the plus three position (Songyang *et al.* 1993). Several reports have described the structure of the SH2 domain with and without bound peptide (Songyang *et al.* 1993; Waksman *et al.* 1993) and suggest that peptide binding produces only localized and relatively small changes to the SH2 domain structure. The SH2 domains of the Src family have been shown to play a role in the regulation of kinase activity by binding the N terminal phosphorylated tyrosine residue and therefore maintaining the kinase in an inactive state.

Patient JP (section 5.3.4) was shown to have a mutation which resulted in the substitution of arginine 307 by glycine. This residue, $\beta$B5 according to the SH2 domain nomenclature, is expected to be found at the base of the phosphotyrosine binding pocket by homology to the Src SH2 domain, and is thought to form two ion pairs with the two of the phosphate oxygens (Waksman *et al.* 1992), as illustrated for the Src SH2 domain in Figure 7.3b. Mutation of this residue to lysine in the oncogene Abl, maintaining the positive charge of the residue, was shown to completely prevent phosphotyrosine binding

(Mayer *et al.* 1992). This residue is the only invariant amino acid in all SH2 domains studied to date, indicating a very strong conservation. Mutation of arginine 337 to glycine in patient JP would therefore be predicted to prevent phosphotyrosine binding to the Btk SH2 domain, so disrupting the function of the protein, and producing the classical XLA phenotype found in this patient.

Patient 276, however, comes from a large pedigree with several affected males, where there is a considerable degree of heterogeneity of phenotype (Mensink *et al.* 1984; de Weers *et al.* 1994b). Some affected boys from this family have classical XLA while others appear to have almost normal levels of immunoglobulin. The mutation in this family is the substitution of arginine 288 by tryptophan. This residue, arginine αA2 according to the nomenclature, is thought to simultaneously recognise the phosphate group, via hydrogen bonds, and the aromatic ring of the tyrosine residue, via an amino aromatic interaction. This could have an important effect on the conformation of the phosphotyrosine in the binding pocket (Waksman *et al.* 1992), illustrated for the Src SH2 domain in Figure 7.3b. Substitution of this residue by tryptophan could be proposed to destabilise the phosphotyrosine/SH2 domain interaction, without completely preventing binding, as this substitution results in the leaky XLA phenotype seen in this family.

The deletion of glycine 302 in patient AC results in a classical XLA phenotype. The exact function of this glycine in the SH2 domain is not known but it is highly conserved in other SH2 domains, suggesting that it has an important role. This residue is found in the AB loop connecting the αA and βB regions, both of which contain residues important for the formation of the phosphotyrosine binding pocket and is found five residues towards the N terminal of the highly conserved arginine βB5. It is possible that this residue has a role in the orientation of the domain to form this pocket.

The SH2 domain amino acid substitution described by Saffran *et al.* ( 1994), tyrosine 361 changed to cysteine, affects the second binding pocket of the domain (residue αB8) which has been shown to bind the residue located at the +3 position after the phosphorylated tyrosine. The three affected brothers described in this report have atypical XLA phenotypes, with B cells up to 2% of peripheral blood lymphocytes and some

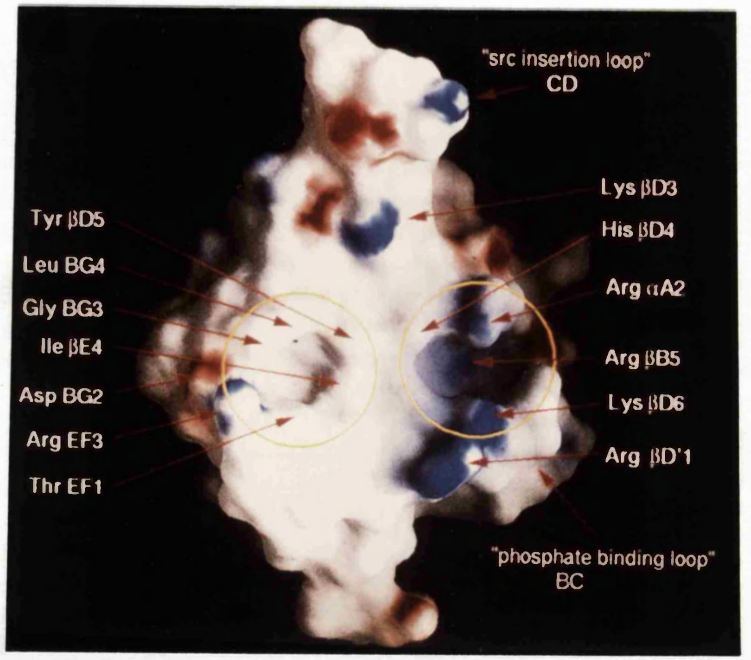
immunoglobulin. Methionine labelling experiments using EBV transformed B cell lines from these patients showed that Btk was produced at normal levels but immunoblotting showed that it had a much shorter half life than normal protein. This result suggests that either this mutation causes conformational instability of the protein, or that the binding of a phosphoprotein to the SH2 domain increases the stability of the protein and this binding is partially prevented by this mutation. A tyrosine to serine substitution has been reported at position 334 in the Btk SH2 domain (Hageman *et al.* 1994), at position βD5 according to the SH2 domain nomenclature. This residue is also thought to be important in the binding of the +3 residue after the phosphotyrosine in the hydrophobic binding pocket, and substitution by serine may alter the interaction of the SH2 domain with the peptide. Mutation analysis of *BTK* has therefore shown the importance of SH2 domain interactions for Btk function.

## 7.11 SH1 domains

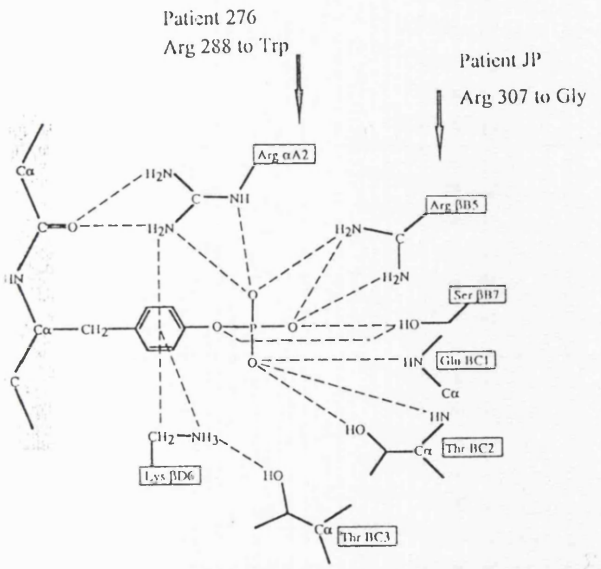
SH1 domains encode the amino acids forming the catalytic domain of the protein which is involved in phosphotransfer, and are reviewed in Hanks (1988). This 250-300 amino acid domain is conserved across the protein kinases and, in most single subunit enzymes, is found at the C terminus of the protein, with the N terminus containing the regulatory sequences. There are regions of high and low conservation throughout all the protein kinases, with the highly conserved regions encompassing the ATP binding and phosphotransfer regions, the autophosphorylation site and the substrate specific sequences. There are short amino acid sequences in the substrate specific domain characterising the serine/threonine kinases and the tyrosine kinases. Within the substrate specific domain of Src related tyrosine kinase proteins the characteristic sequence is HRDLRAAN, as a highly conserved motif. The proteins in the Btk subfamily have the sequence HRDLAARN at this position. The significance of this, if any, is not yet understood.

**Figure 7.3 Representation of the Src SH2 domain**

(a)




(b)




Representations of the Src SH2 domain taken from Waksman *et al.* (1993). (a) The molecular surface of the Src SH2 domain bound to a YEEI peptide (to which the Src SH2 domain specifically binds), with the peptide removed. The surface is coloured according to the local electrostatic potential - deep blue in the most positive regions and deep red in the most negative regions. The two pockets on the surface are outlined in yellow (the phosphotyrosine binding pocket on the left and the +3 residue binding pocket on the right) and important residues are identified by red arrows, including the Arg αA2 residue, altered in patient 276, and Arg βB5, altered in patient JP. (b) Schematic diagram showing the predicted interactions in the phosphotyrosine binding pocket. The amino acid substitutions found in patient 276 and JP are indicated.

In this study, patients FM and TF were both shown to have amino acid substitutions in the kinase domain of Btk. The amino acid alteration in patient FM (section 5.3.2) is within the highly conserved motif responsible for the substrate specificity of the kinase reaction (HR*DLRAAN, with the asterix indicating the altered amino acid). Mutation of arginine 520 to glutamine in this patient would be predicted to disrupt this domain and may potentially result in Btk being unable to specifically phosphorylate proteins on tyrosine. This same mutation has since been described in an additional six individual XLA patients (Zhu *et al.* 1994b; Conley *et al.* 1994; Hageman *et al.* 1994), at least three of which have been shown to be unrelated.

Patient TF and his two brothers all have leaky XLA phenotypes. They were originally diagnosed as XLA on the basis of a non-random pattern of X-inactivation in the B cells of the obligate carrier mother (Alterman *et al.* 1993) and this diagnosis has been confirmed by the identification of a mutation in *BTK*. The mutation results in the substitution of alanine by aspartic acid at residue 607, near the C terminus of the protein (section 5.3.1). This is a non-polar to polar amino acid substitution and, while this is not a highly conserved residue, in Src and other related proteins this residue always appears to be non-polar. The mild phenotype seen in this family suggests that this residue is not in a region critical for the kinase function of the protein, but its substitution by aspartic acid is enough to partially disrupt the kinase activity. Partial disruption of the functioning of Btk in cases such as this will provide valuable insight in to the structure/function relationship of Btk.

Patient AP (section 5.3.1) was shown in this study to have a mutation resulting in a premature termination codon. This would lead to a truncated Btk protein lacking the twenty four amino acids at the C terminal. The severe XLA phenotype seen in this family would suggest that these terminal twenty four amino acids are critical for the stability or functioning of the protein.

Several other kinase domain mutations have been described (Vetrie *et al.* 1993c; Hageman *et al.* 1994; de Weers *et al.* 1994b; Conley *et al.* 1994; Ohta *et al.* 1994; Zhu *et al.* 1994b). The frequent detection of substiution of the two arginines at 520 and 525

in the substrate specific domain as causes of XLA may indicate the relative importance of these residues in the function of the kinase. The R520Q mutation has been reported in six patients in addition to the one reported in this study (patient FM). The R525Q mutation has been reported in three publications, but the unrelatedness of the patients has yet to be established. Ohta et al. (1994) reported that the patients with this mutation have a less severe XLA phenotype, but the patient reported by Vetrie *et al.* (1993) had the classical XLA phenotype. This could suggest that the individual's genetic background and/or previous medical history could be influential on the severity of the XLA defect.

## 7.11 Summary of the spectrum of mutations known to cause XLA

Recent publications have now brought the number of described mutations in *BTK* in XLA patients, including the ones described in this study, to eighty-one (Vetrie *et al.* 1993c; de Weers *et al.* 1994b; Zhu *et al.* 1994b; Zhu *et al.* 1994a; Saffran *et al.* 1994; Ohta *et al.* 1994; Hageman *et al.* 1994; Conley *et al.* 1994; Duriez *et al.* 1994). Several studies have used RT-PCR and DNA sequence analysis to define the mutations (Vetrie *et al.* 1993c; Saffran *et al.* 1994; de Weers *et al.* 1994b), while other studies have used screening methods - SSCP analysis using the *BTK* cDNA (this study) and the *BTK* genomic sequence (Conley *et al.* 1994; Ohta *et al.* 1994), and dideoxy fingerprinting (Zhu *et al.* 1994b). Of these eighty-one mutations, thirty-four are amino acid substitutions (twenty-one different substitutions), sixteen are premature stop codons, three disrupt the initiation of translation codon, fifteen are small insertions or deletions and thirteen result in splicing errors. These results, taken with the approximate 10% of XLA patients who have genomic deletions encompassing all or part of the *BTK* gene, indicate that the mutations which cause XLA are highly heterogeneous. Premature terminations, splicing errors and other nucleotide changes which put translation out of the correct reading frame will all cause major structural changes to the protein leading to lack of function. Amino acid substitutions causing disease, however, provide much more information on the important residues and domains in the protein. If these substitutions

can be correlated with protein function and disease phenotype then much can be learnt about the role of Btk in B cell development.

Six single nucleotide changes causing amino acid substitutions have now been found in more than one patient; R28H (de Weers *et al.* 1994b; Ohta *et al.* 1994; Zhu *et al.* 1994b), T33P (Genevier *et al.* 1994a; Zhu *et al.* 1994b), R520Q (Zhu *et al.* 1994b; Conley *et al.* 1994; Hageman *et al.* 1994; this study), R525Q (Ohta *et al.* 1994; Vetrie *et al.* 1993c; Zhu *et al.* 1994b), R562W (Conley *et al.* 1994; Hageman *et al.* 1994) and M630K (Conley *et al.* 1994; Hageman *et al.* 1994). There appears to be a mutational "hotspot" at amino acid position 520 - seven patients, at least three of whom are reported to be unrelated, have the same mutation (CGA→CAA) which alters the arginine to a glycine this study. Three patients have a different nucleotide alteration at this codon (CGA→TGA) which results in a premature stop codon (Zhu *et al.* 1994b; Hageman *et al.* 1994; Conley *et al.* 1994). Both of these alterations would be caused by the deamination of the methylated cytosine residue in the CpG dinucleotide, resulting in T/G mispairing and the consequent fixing of the mutation at DNA replication.

## 7.12 The role of Btk

Protein phosphorylation is a major mechanism of enzymatic control and signal transduction in the cell and, although phosphorylation of serine and threonine residues make up 95% of the phosphorylated protein in a cell, tyrosine phosphorylation has been shown to play a critical role in signalling pathways from several B cell surface molecules, including the BCR (reviewed in Pleiman *et al.* 1994). Triggering of the antigen specific receptors initiates a series of biochemical cascades, the earliest of which is the phosphorylation of proteins on tyrosine. The BCR does not have intrinsic tyrosine kinase activity and must therefore utilise the activity of recruited cytoplasmic tyrosine kinases. Studies using kinase inhibitors have shown that all known downstream events in signalling and biological responses are dependent on PTK activity.

The importance of Btk in the development of B cells and their immune response has been

unequivocally demonstrated by the identification of mutations in the *BTK* gene in XLA patients, suggesting a uniquely important role for this protein. The definition of this role in B cells, and in the other haematopoietic lineages in which Btk is expressed, however, has yet to be elucidated. The XLA defect appears to be limited to the B cell lineage, as XLA patients appear to have normal macrophage function, and seems to cause a block in B cell differentiation at the pre-B cell stage of development. The expression of the gene from the pro-B cell stage to the mature B cell stage, however, and its absence in fully differentiated plasma cells, suggests that Btk may function at more than one stage of development. As Btk is a non-receptor tyrosine kinase and has such a critical role in the development of B cells, it is highly likely that it functions in one or more B cell signalling pathways.

A number of papers have reported the beginnings of the investigation into Btk function. Cross-linking of surface IgM on B cells has been shown by three laboratories to result in the activation of Btk (de Weers *et al.* 1994a; Saouaf *et al.* 1994) in Daudi and Ramos cells (Burkitt's lymphoma cell lines) and in normal tonsillar B cells. Studies indicated that cross-linking of surface IgM resulted in a significant increase in Btk phosphorylation, predominantly on tyrosine residues, but also on serine and threonine residues (de Weers *et al.* 1994a). An increase in Btk kinase activity was shown using immunoprecipitated Btk, suggesting that the increase in phosphorylation was, at least in part, due to autophosphorylation. This association of Btk with signalling pathways initiated at the BCR suggests a role for Btk in the response to antigenic stimulation of mature B cells. The XLA defect becomes evident, however, at the pre-B cell stage before the expression surface IgM. Pre-B cells express a small amount of $\mu$ on the cell surface in association with surrogate light chains Ig$\alpha$ and Ig$\beta$, as the pre-B cell receptor. The low level of expression of the pre-B cell receptor prevented the analysis of Btk activation on cross-linking of this receptor.

The structural basis by which signals from the BCR are transmitted in B cells has begun to be elucidated and is reviewed in Pleiman *et al.* (1994). The initial reactions are very complex, but have been shown definitively to involve PTK activation. Syk has been shown to associate with the short cytoplasmic tail of the $\mu$ chain of the BCR in resting

B cells. The Src family kinases Fyn, Lyn and Blk are thought to bind to the antigen recognition activation motif (ARAM) of Igα in resting cells via the first ten amino acids of the kinase molecule. On activation of the BCR, the ARAM motifs of Igα and Igβ become phosphorylated on tyrosine and the Src family kinases then bind in a phosphotyrosine dependent manner via their SH2 domains, with the concomitant release of the N terminal binding mechanism. The N terminal region of these kinases is then free to participate in other interactions. The SH3 domains of the Src family kinases Hck, Fyn and Lyn have been shown to bind to the proline rich sequences in the Tec homology domain of Btk. These interactions are specific and do not occur with the SH3 domains of other Src related tyrosine kinases (Cheng *et al.* 1994; Yu *et al.* 1994). This suggests a mechanism by which Btk could associate with Src family kinases that are known to associate with the BCR in resting B cells, linking Btk into the BCR signal transduction pathways.

The cross-linking of the high affinity IgE receptor (Fc$_\epsilon$ RI) on mast cells results in the phosphorylation on tyrosine of several PTKs, including Lyn, Yes, Src and Syk. Kawakami *et al.* (1994) have reported that Btk can be added to this list and that, after cross-linking, a fraction of Btk can be found at the plasma membrane. No direct association with the receptor itself could be detected. The same group have recently shown that the Btk PH domain is constitutively associated with PKC in mast cells (Yao *et al.* 1994). The interaction results in Btk becoming phosphorylated on serine and downregulated. This suggests a role for Btk in a second signal transduction pathway.

Taken together, these results suggest a role for Btk in at least two surface-receptor coupled signal transduction pathways. Btk is probably not constitutively associated with membrane receptors, but may associate with another member of the Src family. However, no critical role has yet been assigned to Btk in pre-B cell development. Expression of the gene throughout B lineage development and in other haematopoietic cell lines, suggests that Btk may be involved in multiple signalling pathways at more than one stage of B cell development. The further study of proteins interacting with Btk in the cell will help in the elucidation of the role of this protein in signal transduction.

## 7.13 Fulfilment of the aims of this study

The aims of this study were as follows:

- To generate a YAC contig across the XLA critical region
- To use cDNA enrichment to isolate the gene responsible for XLA
- Once isolated, to identify mutations in the gene in XLA patients
- To further analyse the gene and the protein and their links with disease

During this study fourteen YACs from the Xq22 region were selected and sized. The generation of YAC insert end probes allowed these YACs to be mapped into a contig of over 2.5Mb spanning the XLA critical region from DXS442 to DXS101. Only one YAC spanned the proximal area of this region, however, and this YAC was later found to have an interstitial deletion. The mapping of this contig allowed the orientation of DNA loci in the region and facilitated the selection of a YAC to use to identify candidate genes for XLA. Magnetic bead technologies were used to immobilise YAC DNA and cDNA enrichment was used to identify transcribed sequences from the region. After the gene responsible for XLA was identified (Vetrie *et al.* 1993c; Tsukada *et al.* 1993), eighteen XLA patients were analysed for mutations. Fifteen mutations were identified including splice site mutations, premature stop codons, small insertions and deletions, amino acid substitutions and disruptions of the initiation codon. Some of these mutations were found to be in the predicted functional domains of the protein (in the PH domain, the SH2 domain and the kinase domain) permitting speculation on the functional effects of the mutations. *BTK* mutation analysis was shown to result in improved carrier determination in families affected by XLA. The involvement of *BTK* in other disease phenotypes, including "*xid*-like" disease, hypogammaglobulinemia and B cell leukaemia, was assessed but no evidence was found to suggest any association. An initial study of the mutant Btk protein from three XLA patients showed, in one case, that a mutation in the SH2 domain of the protein does not appear to affect kinase function and that in two cases, the protein appeared to be unstable.

## 7.14 Future work

The mutations in *BTK* presented here provide the starting point of an investigation into Btk protein function and the defects which cause XLA. The premature terminations and mutations which result in the Btk protein being translated in the wrong reading frame will produce gross changes in the protein structure, leaving it without function. The amino acid substitutions and other changes which leave the protein in frame, however, provide much more information into the structure/function relationship of the protein. Further studies of these mutant proteins, in the form of expression and stability studies and *in vitro* kinase assays, will further clarify the postulated basis of the disease in these patients. Further analysis of *BTK* from the patients with the leaky XLA phenotypes will be especially important in the elucidation of the function of Btk in the B cell. A reduction in kinase activity or binding of another protein resulting from a specific amino acid substitution furthers our knowledge of the residues involved in these processes.

It has already been established that Btk functions in one or several B cell signalling pathways. To fully elucidate the role of Btk in these pathways, it will be important to identify proteins which interact with Btk. Studies are underway in this laboratory to identify such proteins using glutathione-S-transferase fusion proteins containing the N terminal domain, the SH3 domain and the SH2 domain of Btk. These experiments could also be approached using the yeast two hybrid system (reviewed in Fields & Sternglanz, 1994). This technique is based on the creation of two fusion proteins; one with the protein of interest fused with a DNA binding domain of a transcriptional activator, and the other created from cDNA containing genes which may encode proteins interacting with the target protein library fused to the activation domain of a transcriptional activator. Interacting proteins will then result in the activation of transcription of a reporter gene. Isolation of proteins interacting with Btk will be an important part of the elucidation of the role of Btk. The specific binding of proteins to the different domains of Btk may be prevented by the use of fusion proteins containing amino acid substitutions found in Btk in XLA patients. This may then implicate certain residues in protein-protein interactions. GST-SH2 domain fusion proteins containing the amino acid substitutions found in patients JP and 276 (section 5.3.4) have already been constructed (G.Cory and R.Lovering,

personal communication). Identification of the critical interactions involving Btk occurring in the cell will further elucidate the role of Btk in the pathogenesis of XLA.

Over-expression of Btk and the expression of mutant Btk proteins in cell lines may provide a useful tool for studying the function of Btk. The study of Btk activity, the pattern of phosphorylated proteins found after the stimulation of different cell surface molecules and the identification of proteins interacting with Btk in these cell lines will then add to knowledge on the molecular basis of XLA.

Btk knock-out mice may help to further elucidate the relationship between human XLA and mouse *xid*. Mice with deletions of part of the PH domain and of the ATP binding site have already been generated (P.Sideras, personal communication). The phenotypic consequences of these deletions are still being assessed. The differences in phenotype between XLA and *xid* appear to be due to different mechanisms of B cell ontogeny in humans and mice, respectively, as very similar amino acid substitutions at the same residue produce different phenotypes in each species. It appears, therefore, that the mouse is not a good animal model of XLA.

## 7.15 Implications for the treatment and potential cure of XLA

Fast and reliable carrier detection will be available to some women who may be carriers of XLA as a result of mutation detection in the *BTK* gene. This will allow genetic counselling to be based on more accurate information. Pre-natal diagnosis and early disease detection will allow administration of prophylactic antibiotics and intravenous immunoglobulin before the patient suffers from infection. The treatment for XLA, however, will not change due to the identification of the gene. In the future, the idea of gene therapy as a cure for XLA is very appealing. The insertion of a correct copy of the gene into stem cells in the bone marrow has already been attempted for adenosine deaminase deficiency, and is theoretically possible for XLA. It must first be established, however, that over expression of Btk will not result in the unregulated growth or differentiation of B cells. Studies of leukaemia patients have not yet shown any link with

Btk over-expression or mutations in the gene. It must also be established that mutant Btk proteins do not act in the dominant-negative fashion seen for some other proteins (Herskowitz, 1987). In these cases, the correct function of an introduced protein is over-ridden by the presence of a mutant protein. Gene therapy for XLA remains in the future, but once the gene therapy technologies are fully established, it may become a reality.

# REFERENCES


Ackerman, S.K. & Douglas, S.D. (1978). Purification of human monocytes on microexudate-coated sufaces. *J. Immunol.* **120**, 1372-1374.

Albertsen, H.M., Abderrahim, H., Cann, H.M., Dausset, J., Le Paslier, D. & Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4256-4260.

Allen, R.C., Armitage, R.J., Conley, M.E., Rosenblatt, H., Jenkins, N.A., Copeland, N.G., Bedell, M.A., Edelhoff, S., Disteche, C.M., Simoneaux, D.K., Fanslow, W.C., Belmont, J., Spriggs, M.K. (1993). CD40 ligand gene defects responsible for X-linked hyper-IgM syndrome. *Science* **259**, 990-993.

Allen, R.C. & Belmont, J.W. (1992). Dinucleotide repeat polymorphism at the DXS178 locus. *Hum. Mol. Genet.* **1**, 216.

Alterman, L.A., de Alwis, M., Genet, S., Lovering, R., Middleton Price, H., Morgan, G., Jones, A., Malcolm, S., Levinsky, R.J. & Kinnon, C. (1993). Carrier determination for X-linked agammaglobulinemia using X inactivation analysis of purified B cells. *J. Immunol. Methods* **166**, 111-116.

Ambrosino, D.M., Siber, G.R., Chilmonczyk, B.A., Jernberg, J.B. & Finberg, R.W. (1987). An immunodeficiency characterized by impaired antibody responses to polysaccharides. *N. Engl. J. Med.* **316**, 790-793.

Anand, R., Villasante, A. & Tyler Smith, C. (1989). Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed-field gel electrophoresis. *Nucleic Acids Res.* **17**, 3425-3433.

Anand, R., Riley, J.H., Butler, R., Smith, J.C. & Markham, A.F. (1990). A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage. *Nucleic Acids Res.* **18**, 1951-1956.

Anker, R., Conley, M.E. & Pollok, B.A. (1989). Clonal diversity in the B cell repertoire of patients with X-linked agammaglobulinemia. *J. Exp. Med.* **169**, 2109-2119.

Aruffo, A., Farrington, M., Hollenbaugh, D., Li, X., Milatovich, A., Nonoyama, S., Bajorath, J., Grosmaire, L.S., Stenkamp, R., Neubauer, M., Roberts, R.L., Ledbetter, J.A., Francke, U. and Ochs, H.D. (1993). The CD40 ligand, gp39, is defective in activated T cells from patients with X-linked hyper-IgM syndrome. *Cell* **72**, 291-300.

Arveiler, B., Oberle, I. & Mandel, J.L. (1987). Genetic mapping of nine DNA markers in the q11-q22 region of the human X chromosome. *Genomics* **1**, 60-66.

Ballabio, A., Carrozzo, R., Parenti, G., Gil, A., Zollo, M., Persico, M.G., Gillard, E.,

Affara, N., Yates, J., Ferguson Smith, M.A., Frants, R.R., Eriksson, A.W. & Andria, G. (1989). Molecular heterogeneity of steroid sulfatase deficiency: a multicenter study on 57 unrelated patients, at DNA and protein levels. *Genomics* **4**, 36-40.

Barker, D.F., Fain, P.R., Goldgar, D.E., Dietz Band, J.N., Turco, A.E., Kashtan, C.E., Gregory, M.C., Tryggvason, K., Skolnick, M.H. & Atkin, C.L. (1991). High-density genetic and physical mapping of DNA markers near the X-linked Alport syndrome locus: definition and use of flanking polymorphic markers. *Hum. Genet.* **88**, 189-194.

Bates, G. & Lehrach, H. (1994). Trinucleotide repeat expansions and human genetic disease. *Bioessays* **16**, 277-284.

Bell, L.R., Maine, E.M., Schedl, P. & Cline, T.W. (1988). *Sex-lethal*, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding protein. *Cell* **55**, 1037-1046.

Bentley, D.R., Todd, C., Collins, J., Holland, J., Dunham, I., Hassock, S., Bankier, A. & Giannelli, F. (1992). The development and application of automated gridding for efficient screening of yeast and bacterial ordered libraries. *Genomics* **12**, 534-541.

Bernstein, H.S., Bishop, D.F., Astrin, K.H., Kornreich, R., Eng, C.M., Sakuraba, H. & Desnick, R.J. (1989). Fabry disease: six gene rearrangements and an exonic point mutation in the alpha-galactosidase gene. *J. Clin. Invest.* **83**, 1390-1399.

Bickmore, W.A. & Bird, A.P. (1992). Use of restriction endonucleases to detect and isolate genes from mammalian cells. *Methods Enzymol.* **216**, 224-244.

Bird, A.P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends. Genet.* **3**, 342-347.

Bleeker-Wagemakers, E., Orth, U., Meire, F., Schwinger, E. & Gal, A. (1991). Gene for X-linked megalocornea is closely linked to loci in Xq21-q22. *Cytogenet. Cell Genet.* **58**, A26852(Abstract)

Boyd, Y. & Fraser, N.J. (1990). Methylation patterns at the hypervariable X-chromosome locus DXS255 (M27 beta): correlation with X-inactivation status. *Genomics* **7**, 182-187.

Brown, T.C. & Jiricny, J. (1987). A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* **50**, 945-950.

Brownstein, B.H., Silverman, G.A., Little, R.D., Burke, D.T., Korsmeyer, S.J., Schlessinger, D. & Olson, M.V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244**, 1348-1351.

Bruns, G., Gusella, J., Keys, C., Leary, A., Housman, D. & Gerald, P. (1982).

Isolation of X chromosme DNA sequences. *Adv. Exp. Med. Biol.* **154**, 60-72.

Bruton, O.C. (1952). Agammaglobulinemia. *Pediatrics* **9**, 722-727.

Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A. & Housman, D.E. (1991). Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 4005-4009.

Buluwela, L., Forster, A., Boehm, T. & Rabbitts, T.H. (1989). A rapid procedure for colony screening using nylon filters. *Nucleic Acids Res.* **17**, 452.

Burke, D.T., Carle, G.F. & Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-812.

Buzi, F., Notarangelo, L.D., Plebani, A., Duse, M., Parolini, O., Monteleone, M. & Ugazio, A.G. (1994). X-linked agammaglobulinemia, growth hormone deficiency and delay of growth and puberty. *Acta Paediatr.* **83**, 99-102.

Cambier, J.C., Pleiman, C.M. & Clark, M.R. (1994). Signal transduction by the B cell receptor and its coreceptors. *Annu. Rev. Immunol.* **12**, 457-486.

Campana, D., Farrant, J., Inamdar, N., Webster, A.D. & Janossy, G. (1990). Phenotypic features and proliferative activity of B cell progenitors in X-linked agammaglobulinemia. *J. Immunol.* **145**, 1675-1680.

Cheng, G., Ye, Z.-S. & Baltimore, D. (1994). Binding of Bruton's tyrosine kinase to Fyn, Lyn or Hck through a Src homolgy 3 domain-mediated interaction. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 8152-8155.

Cheng, J., Fogel Petrovic, M. & Maquat, L.E. (1990). Translation to near the distal end of the penultimate exon is required for normal levels of spliced triosephosphate isomerase mRNA. *Mol. Cell. Biol.* **10**, 5215-5225.

Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159.

Chu, G., Vollrath, D. & Davis, R.W. (1986). Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* **234**, 1582-1585.

Chumakov, I.M., Le Gall, I., Billault, A., Ougen, P., Soularue, P., Guillou, S., Rigault, P., Bui, H., De Tand, M.F., Barillot, E., Abderrahim, H., Cherif, D., Berger, R., Le Paslier, D. & Cohen, D. (1992). Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nat. Genet.* **1**, 222-225.

Cole, C.G., Goodfellow, P.N., Bobrow, M. & Bentley, D.R. (1991). Generation of novel sequence tagged sites (STSs) from discrete chromosomal regions using Alu-PCR

[published erratum appears in Genomics 1991 Nov;11(3):782]. *Genomics* **10**, 816-826.

Conley, M.E. (1985). B Cells in patients with X-linked agammaglobulinemia. *J. Immunol.* **134**, 3070

Conley, M.E., Brown, P., Pickard, A.R., Buckley, R.H., Miller, D.S., Raskind, W.H., Singer, J.W. & Fialkow, P.J. (1986). Expression of the gene defect in X-linked agammaglobulinemia. *New Eng. J. Med.* **315**, 564

Conley, M.E., Sullivan, J.L., Neidich, J.A. & Puck, J.M. (1990). X chromosome inactivation patterns in obligate carriers of X-linked lymphoproliferative syndrome. *Clin. Immunol. Immunopathol.* **55**, 486-491.

Conley, M.E., Burks, A.W., Herrod, H.G. & Puck, J.M. (1991). Molecular analysis of X-linked agammaglobulinemia with growth hormone deficiency. *J. Pediatr.* **119**, 392-397.

Conley, M.E., Fitch-Hilgenberg, M.E., Cleveland, J.L., Parolini, O. & Rohrer, J. (1994). Screening of genomic DNA to identify mutations in the gene for Bruton's tyrosine kinase. *Hum. Mol. Genet.* **3**, 1751-1756.

Conley, M.E. & Puck, J.M. (1988). Carrier detection in typical and atypical X-linked agammaglobulinemia. *J. Pediatr.* **112**, 688-694.

Conley, M.E. & Sweinberg, S.K. (1992). Females with a disorder phenotypically identical to X-linked agammaglobulinemia. *J. Clin. Immunol.* **12**, 139-143.

Cooper, D.N. & Krawczak, M. (1991). Mechanisms of insertional mutagenesis in human genes causing genetic disease. *Hum. Genet.* **87**, 409-415.

Cooper, D.N. & Youssoufian, H. (1988). The CpG dinucleotide and human genetic disease. *Hum. Genet.* **78**, 151-155.

Cotton, R.G., Rodrigues, N.R. & Campbell, R.D. (1988). Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4397-4401.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S. & Myers, R.M. (1990). Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**, 245-250.

Craig, J.M. & Bickmore, W.A. (1994). The distribution of CpG islands in mammlian chromosomes. *Nat. Genet.* **7**, 376-382.

Crampton, J.M., Davies, K.E. & Knapp, T.F. (1981). The occurence of families of repetitive sequnces in a library of cloned cDNA from human lymphocytes. *Nuc. Acid. Res.* **9**, 3821-3834.

Cremers, F.P., van de Pol, D.J., van Kerkhoff, L.P., Wieringa, B. & Ropers, H.H. (1990). Cloning of a gene that is rearranged in patients with choroideraemia. *Nature* **347**, 674-677.

Danenberg, P.V., Horikoshi, T., Volkenandt, M., Danenberg, K., Lenz, H.J., Shea, L.C., Dicker, A.P., Simoneau, A., Jones, P.A. & Bertino, J.R. (1992). Detection of point mutations in human DNA by analysis of RNA conformation polymorphism(s). *Nucleic Acids Res.* **20**, 573-579.

de Weers, M., Verschuren, M.C., Kraakman, M.E., Mensink, R.G., Schuurman, R.K., van Dongen, J.J. & Hendriks, R.W. (1993). The Bruton's tyrosine kinase gene is expressed throughout B cell differentiation, from early precursor B cell stages preceding immunoglobulin gene rearrangement up to mature B cell stages. *Eur. J. Immunol.* **23**, 3109-3114.

de Weers, M., Brouns, G.S., Hinshelwood, S., Kinnon, C., Schuurman, R.K.B., Hendriks, R.W. & Borst, J. (1994a). Tyrosine phosphorylation and activation of the human Bruton's tyrosine kinase in response to B-cell receptor stimulation. *J. Biol. Chem.* **269**, 23857-23860.

de Weers, M., Mensink, R.G., Kraakman, M.E., Schuurman, R.K. & Hendriks, R.W. (1994b). Mutation analysis of the Bruton's tyrosine kinase gene in X-linked agammaglobulinemia: identification of a mutation which affects the same codon as is altered in immunodeficient mice. *Hum. Mol. Genet.* **3**, 161.

Derry, J.M., Ochs, H.D. & Francke, U. (1994). Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. *Cell* **78**, 635-644.

Dietz Band, J.N., Turco, A.E., Willard, H.F., Vincent, A., Skolnick, M.H. & Barker, D.F. (1990). Isolation, characterization, and physical localization of 33 human X-chromosome RFLP markers. *Cytogenet. Cell Genet.* **54**, 137-141.

Dietz, H.C., Valle, D., Francomano, C.A., Kendzior, R.J., Jr., Pyeritz, R.E. & Cutting, G.R. (1993). The skipping of constitutive exons in vivo induced by nonsense mutations. *Science* **259**, 680-683.

Dietz, H.C. & Kendzior, R.J. (1994). Maintenance of an open reading frame as an additional level of scrutiny during splice site selection. *Nat. Genet.* **8**, 183-188.

DiSanto, J.P., Bonnefoy, J.Y., Gauchat, J.F., Fischer, A. & de Saint Basile, G. (1993). CD40 ligand mutations in x-linked immunodeficiency with hyper-IgM. *Nature* **361**, 541-543.

Dryja, T.P., McGee, T.L., Reichel, E., Hahn, L.B., Cowley, G.S., Yandell, D.W., Sandberg, M.A. & Berson, E.L. (1990). A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature* **343**, 364-366.

Duriez, B., Duquesnoy, P., Dastot, F., Bougneres, P., Amselem, S. & Goossens, M. (1994). An exon-skipping mutation in the btk gene of a patient with X-linked agammaglobulinemia and isolated growth hormone deficiency. *FEBS Lett.* **346**, 165-170.

Duyk, G.M., Kim, S.W., Myers, R.M. & Cox, D.R. (1990). Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8995-8999.

Echols, H. & Goodman, M.F. (1991). Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* **60**, 477-511.

Eck, M.J., Shoelson, S.E. & Harrison, S.C. (1993). Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature* **362**, 87-91.

Ehrlich, M. & Wang, Y.-H. (1981). 5-Methylcytosine in eukaryotic DNA. *Science* **212**, 1350-1357.

Elvin, P., Slynn, G., Black, D., Graham, A., Butler, R., Riley, J., Anand, R. & Markham, A.F. (1990). Isolation of cDNA clones using yeast artificial chromosome probes. *Nucleic Acids Res.* **18**, 3913-3917.

Eperon, I.C., Ireland, D.C., Smith, R.A., Mayeda, A. & Krainer, A.R. (1993). Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *EMBO J.* **12**, 3607-3617.

Fan, E., Levin, D.B., Glickman, B.W. & Logan, D.M. (1993). Limitations in the use of SSCP analysis. *Mutat. Res.* **288**, 85-92.

Fan, W.F., Wei, X., Shukla, H., Parimoo, S., Xu, H., Sankhavaram, P., Li, Z. & Weissman, S.M. (1993). Application of cDNA selection techniques to regions of the human MHC. *Genomics* **17**, 575-581.

Fearon, E.R., Winkelstein, J.A., Civin, C.I., Pardoll, D.M. & Vogelstein, B. (1987). Carrier detection in X-linked agammaglobulinemia by analysis of X-chromosome inactivation. *N. Engl. J. Med.* **316**, 427-431.

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

Feinberg, A.P. & Vogelstein, B.A. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity: addendum. *Anal. Biochem.* **137**, 266-267.

Fields, C., Adams, M.D., White, O. & Venter, J.C. (1994). How many genes in the human genome? *Nature Genetics* **7**, 345-236.

Fields, S. & Sternglanz, R. (1994). The two-hybrid system: an assay for protein-protein interactions. *Trends. Genet.* **10**, 286-292.

Fleischer, T.A., White, R.M., Broder, S., Nissley, P., Blaese, R.M., Mulvihill, J.J., Olive, G. & Waldmann, T.A. (1980). X-linked hypogammaglobulinemia and isolated growth hormone deficiency. *N. Engl. J. Med.* **302**, 1429-1433.

Foote, S., Vollrath, D., Hilton, A. & Page, D.C. (1992). The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science* **258**, 60-66.

Forrest, S.M., Dahl, H.H., Howells, D.W., Dianzani, I. & Cotton, R.G. (1991). Mutation detection in phenylketonuria by using chemical cleavage of mismatch: importance of using probes from both normal and patient samples [published erratum appears in Am J Hum Genet 1992 Mar;50(3):659]. *Am. J. Hum. Genet.* **49**, 175-183.

Fountain, J.W., Wallace, M.R., Bruce, M.A., Seizinger, B.R., Menon, A.G., Gusella, J.F., Michels, V.V., Schmidt, M.A., Dewald, G.W. & Collins, F.S. (1989). Physical mapping of a translocation breakpoint in neurofibromatosis. *Science* **244**, 1085-1087.

Genevier, H.C., Hinshelwood, S., Gaspar, H.B., Rigley, K.P., Brown, D., Saeland, S., Rousset, F., Levinsky, R.J., Callard, R.E., Kinnon, C. & Lovering, R.C. (1994). Expression of Bruton's tyrosine kinase protein within the B cell lineage. *Eur. J. Immunol.* (In Press)

Gibson, T., Hyvonen, M., Musacchio, A., Saraste, M. & Birney, E. (1994). PH domain: the first anniversary. *TIBS* **19**, 349-353.

Goldblum, R.M., Lord, R.A., Cooper, M.D., Gathings, W.E. & Goldman, A.S. (1974). X-linked B lymphocyte deficiency. *J. Pediatr.* **85**, 188-191.

Goonewardena, P., Sjoholm, A.G., Nilsson, L.-A. & Pettersson, U. (1988). Linkage analysis of the properdin deficiency gene: suggestion of a locus in the proximal part of the short arm of the X chromosome. *Genomics* **2**, 115-118.

Goundis, D., Holt, S.M., Boyd, Y. & Reid, K.B. (1989). Localization of the properdin structural locus to Xp11.23-Xp21.1. *Genomics* **5**, 56-60.

Gout, I., Dhand, R., Hiles, I.D., Fry, M.J., Panayotou, G., Das, P., Truong, O., Totty, N.F., Hsuan, J., Booker, G.W., Campbell, I.D. and Waterfield, M. (1993). The GTPase dynamin binds to and is activated by a subset of SH3 domains. *Cell* **75**, 25-36.

Greaves, M.F. (1988). Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia* **2**, 120-125.

Green, E.D., Riethman, H.C., Dutchik, J.E. & Olson, M.V. (1991). Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* **11**, 658-669.

Green, E.D. & Olson, M.V. (1990). Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 1213-1217.

Grierson, H. & Purtilo, D.T. (1987). Epstein-Barr virus infections in males with the X-linked lymphoproliferative syndrome. *Ann. Intern. Med.* **106**, 538-545.

Guioli, S., Arveiler, B., Bardoni, B., Notarangelo, L.D., Panina, P., Duse, M., Ugazio, A., Oberle, I., de Saint Basile, G., Mandel, J.L. and Camerino, G. (1989). Close linkage of probe p212 (DXS178) to X-linked agammaglobulinemia. *Hum. Genet.* **84**, 19-21.

Hageman, T.L., Chen, Y., Rosen, F.S. & Kwan, S.-P. (1994). Genomic organisation of the Btk gene and exon scanning for mutations in patients with X-linked agammaglobulinemia. *Hum. Mol. Genet.* **3**, 1743-1749.

Hanks, S.K., Quinn, A.M. & Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42-52.

Harlan, J.E., Hajduk, P.J., Yoon, H.S. & Fesik, S.W. (1994). Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate. *Nature* **371**, 168-170.

Hartley, D.A., Davies, K.E., Drayna, D., White, R.L. & Williamson, R. (1984). A cytological map of the human X chromosome - evidence for non-random recombination. *Nucleic Acid. Res.* **12**, 5277-5285.

Haslam, R.J., Koide, H.B. & Hemmings, B.A. (1993). Pleckstrin domain homology [letter]. *Nature* **363**, 309-310.

Hayward, C., Rae, A.L., Porteous, M.E., Logie, L.J. & Brock, D.J. (1994). Two novel mutations and a neutral polymorphism in EGF-like domains of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan syndrome patients. *Hum. Mol. Genet.* **3**, 373-375.

Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G. & Galas, D.J. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* **17**, 6545-6551.

Hendriks, R.W., Mensink, E.J., Kraakman, M.E., Thompson, A. & Schuurman, R.K. (1989). Evidence for male X chromosomal mosaicism in X-linked agammaglobulinemia. *Hum. Genet.* **83**, 267-270.

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *Nature* **329**, 219-222.

Hirai, H. & Varmus, H.E. (1990). Site-directed mutagenesis of the SH2- and SH3-coding domains of c-src produces varied phenotypes, including oncogenic activation of p60c-src.

*Mol. Cell. Biol.* **10**, 1307-1318.

Hofker, M.H., Bergen, A.A., Skraastad, M.I., Carpenter, N.J., Veenema, H., Connor, J.M., Bakker, E., van Ommen, G.J. & Pearson, P.L. (1987). Efficient isolation of X chromosome-specific single-copy probes from a cosmid library of a human X/hamster hybrid-cell line: mapping of new probes close to the locus for X-linked mental retardation. *Am. J. Hum. Genet.* **40**, 312-328.

Hongyo, T., Buzard, G.S., Calvert, R.J. & Weghorst, C.M. (1993). 'Cold SSCP': a simple, rapid and non-radioactive method for optimized single-strand conformation polymorphism analyses. *Nucleic Acids Res.* **21**, 3637-3642.

Iwahana, H., Yoshimoto, K., Mizusawa, N., Kudo, E. & Itakura, M. (1994). Multiple fluorescence-based PCR-SSCP analysis. *Biotechniques* **16**, 296-7, 300.

Jones, C.P., Janson, M. & Nodenskjold, M. (1989). Separation of yeast chromosomes in the megabase range suitable as size markers for pulsed field electrophoresis. *Technique* **1**, 90-95.

Katz, F.E., Lovering, R.C., Bradley, L.A., Rigley, K.P., Brown, D., Cotter, F., Chessells, J.M., Levinsky, R.J. & Kinnon, C. (1994). Expression of the X-linked agammaglobulinemia gene, btk in B-cell acute lymphoblastic leukemia. *Leukemia* **8**, 574-577.

Kawahara, A., Minami, Y. & Taniguchi, T. (1994). Evidence for a critical role for the cytoplasmic region of the interleukin 2 (IL-2) receptor gamma chain in IL-2, IL-4 and IL-7 signalling. *Mol. Cell. Biol.* **14**, 5433-5440.

Kawakami, Y., Yao, L., Miura, T., Tsukada, S., Witte, O.N. & Kawakami, T. (1994). Tyrosine phosphorylation and activation of Bruton tyrosine kinase upon Fc epsilon RI cross-linking. *Mol. Cell. Biol.* **14**, 5108-5113.

Keen, J., Lester, D., Inglehearn, C., Curtis, A. & Bhattacharya, S. (1991). Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. *Trends. Genet.* **7**, 5.

Korn, B., Sedlacek, Z., Manca, A., Kioschis, P., Konecki, D., Lehrach, H. & Poustka, A. (1992). A strategy for the selection of transcribed sequences in the Xq28 region. *Hum. Mol. Genet.* **1**, 235-242.

Korthauer, U., Graf, D., Mages, H.W., Briere, F., Padayachee, M., Malcolm, S., Ugazio, A.G., Notarangelo, L.D., Levinsky, R.J. & Kroczek, R.A. (1993). Defective expression of T-cell CD40 ligand causes X-linked immunodeficiency with hyper-IgM. *Nature* **361**, 539-541.

Kouprina, N., Eldarov, M., Moyzis, R., Resnick, M. & Larionov, V. (1994). A model system to assess the integrity of mammalian YACs during transformation and propagation

in yeast. *Genomics* **21**, 7-17.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125-8148.

Krawczak, M., Reiss, J. & Cooper, D.N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* **90**, 41-54.

Krawczak, M. & Cooper, D.N. (1991). Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum. Genet.* **86**, 425-441.

Kroczek, R.A., Graf, D., Brugnoni, D., Giliani, S., Korthauer, U., Ugazio, A., Senger, G., Mages, H.W., Villa, A. & Notarangelo, L.D. (1994). Defective expression of CD40 ligand on T cells causes "X-linked immunodeficiency with hyper IgM (HIGM1)". *Immunol. Rev.* **138**, 39-59.

Kurzrock, R., Gutterman, J.U. & Talpaz, M. (1988). The molecular genetics of Philadelphia chromosome-positive leukemias. *N. Engl. J. Med.* **319**, 990-998.

Kwan, S.-P., Kunkel, L., Bruns, G., Wedgewood, R.J., Latt, S. & Rosen, F.S. (1986). Mapping of the X-linked agammaglobulinemia locus by use of restriction fragment-length polymorphism. *J. Clin. Invest.* **77**, 649-652.

Kwan, S.-P., Terwilliger, J., Hagemann, T., Ott, J., Ochs, H., Wedgwood, R. & Rosen, F. (1991). Linkage mapping of closely linked DNA markers for X-linked agammglobulinemia. *Cytogenet. Cell Genet.* **58**, A27489 (Abstract).

Kwan, S.P., Terwilliger, J., Parmley, R., Raghu, G., Sandkuyl, L.A., Ott, J., Ochs, H., Wedgwood, R. & Rosen, F. (1990). Identification of a closely linked DNA marker, DXS178, to further refine the X-linked agammaglobulinemia locus. *Genomics* **6**, 238-242.

Lakich, D., Kazazian, H.H., Jr., Antonarakis, S.E. & Gitschier, J. (1993). Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat. Genet.* **5**, 236-241.

Lamond, A.I. (1993). The spliceosome. *Bioessays* **15**, 595-603.

Langlois, S., Kastelein, J.J. & Hayden, M.R. (1988). Characterization of six partial deletions in the low-density-lipoprotein (LDL) receptor gene causing familial hypercholesterolemia (FH). *Am. J. Hum. Genet.* **43**, 60-68.

Larin, Z., Monaco, A.P. & Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 4123-4127.

Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics* **13**, 1095-1107.

Lathrop, G.M. & Laloulel, J.M. (1988). Efficient computations in multilocus linkage analysis. *Am. J. Hum. Genet.* **42**, 498-505.

Lau, Y.L., Levinsky, R.J., Malcolm, S., Goodship, J., Winter, R. & Pembrey, M. (1988). Genetic prediction in X-linked agammaglobulinaemia. *Am. J. Med. Genet.* **31**, 437-448.

Lee, J.T., Murgia, A., Sosnoski, D.M., Olivos, I.M. & Nussbaum, R.L. (1992). Construction and characterization of a yeast artificial chromosome library for Xpter-Xq27.3: a systematic determination of cocloning rate and X-chromosome representation. *Genomics* **12**, 526-533.

Leonard, W.J., Noguchi, M., Russell, S.M. & McBride, O.W. (1994). The molecular basis of X-linked severe combined immunodeficiency: the role of the interleukin-2 receptor gamma chain as a common gamma chain. *Immunol. Rev.* **138**, 61-86.

Lewin, B. (1993). *Genes V.* Wiley New York,

Lindsay, S. & Bird, A.P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* **327**, 336-338.

Lindsberg, M.L., Brunswick, M., Yamada, H., Lees, A., Inman, J., June, C.H. & Mond, J.J. (1991). Biochemical analysis of the immune B cell defect in xid mice. *J. Immunol.* **147**, 3774-3779.

Little, R.D., Pilia, G., Johnson, S., D'Urso, M. & Schlessinger, D. (1992). Yeast artificial chromosomes spanning 8 megabases and 10-15 centimorgans of human cytogenetic band Xq26. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 177-181.

Liu, X., Brodeur, S.R., Gish, G., Zhou, S., Cantley, L.C., Laudano, A.P. & Pawson, T. (1993). Regulation of c-Src tyrosine kinase activity by the Src SH2 domain. *Oncogene* **8**, 1119-1126.

Lovering, R., Middleton Price, H.R., O'Reilly, M.A., Genet, S.A., Parkar, M., Sweatman, A.K., Bradley, L.D., Alterman, L.A., Malcolm, S., Morgan, G., Levinsky, R.J. and Kinnon, C. (1993a). Genetic linkage analysis identifies new proximal and distal flanking markers for the X-linked agammaglobulinemia gene locus, refining its localization in Xq22. *Hum. Mol. Genet.* **2**, 139-141.

Lovering, R., Sweatman, A.K., O'Reilly, M.A., Genet, S.A., Middleton Price, H., Malcolm, S., Levinsky, R.J. & Kinnon, C. (1993b). Physical mapping identifies DXS265 as a useful genetic marker for carrier detection and prenatal diagnosis of X-linked agammaglobulinemia. *Hum. Genet.* **91**, 178-180.

Lovering, R.C., Sweatman, A.K., Genet, S.A., Middleton-Price, H., Vetrie, D., Vorechovsky, I., Bentley, D., Fontan, G., Espanol, T., Morgan, G., Levinsky, R. & Kinnon, C. (1994). Identification of deletions in the *btk* gene allows unambiguous assessment of carrier status in families with X-linked agammaglobulinemia. *Hum. Genet.* **94**, 77-79.

Lovett, M., Kere, J. & Hinton, L.M. (1991). Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9628-9632.

Macias, M.J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M. & Oschkinat, H. (1994). Structure of the pleckstrin homology domain from beta-spectrin. *Nature* **369**, 675-677.

Malcolm, S., de Saint Basile, G., Arveiler, B., Lau, Y.L., Szabo, P., Fischer, A., Griscelli, C., Debre, M., Mandel, J.L., Callard, R.E., Robertson, M.E., Goodship, J.E., Pembrey, M.E. and Levinsky, R.J. (1987). Close linkage of random DNA fragments from Xq 21.3-22 to X-linked agammaglobulinaemia (XLA). *Hum. Genet.* **77**, 172-174.

Mano, H., Mano, K., Tang, B., Koehler, M., Yi, T., Gilbert, D.J., Jenkins, N.A., Copeland, N.G. & Ihle, J.N. (1993). Expression of a novel form of Tec kinase in hematopoietic cells and mapping of the gene to chromosome 5 near Kit. *Oncogene* **8**, 417-424.

Manoni, M., Tribioli, C., Lazzari, B., DeBellis, G., Patrosso, C., Pergolizzi, R., Pellegrini, M., Maestrini, E., Rivella, S., Vezzoni, P. & Toniolo, D. (1991). The nucleotide sequence of a CpG island demonstrates the presence of the first exon of the gene encoding the human lysosomal membrane protein lamp2 and assigns the gene to Xq24. *Genomics* **9**, 551-554.

Mayer, B.J., Jackson, P.K., Van Etten, R.A. & Baltimore, D. (1992). Point mutations in the abl SH2 domain coordinately impair phosphotyrosine binding in vitro and transforming activity in vivo. *Mol. Cell. Biol.* **12**, 609-618.

Mayer, B.J., Ren, R., Clark, K.L. & Baltimore, D. (1993). A putative modular domain present in diverse signaling proteins [letter]. *Cell* **73**, 629-630.

Mayer, B.J. & Baltimore, D. (1994). Signalling through SH2 and SH3 domains. *Trends Cell Biol.* **3**, 8-13.

Mensink, E.J.B.M., Schot, J.D.L., Tippett, P. & Schuurman, R.K.B. (1984). X-linked agammaglobulinemia and the red blood cell determinants Xg and 12E7 are not closely linked. *Hum. Genet.* **68**, 303-309.

Mensink, E.J.B.M., Thompson, A., Schot, J.D.L., van de Greef, W.M.M., Sandkuyl,

L.A. & Schuurman, R.K.B. (1986). Mapping of a gene for X-linked agammaglobulinemia and evidence for genetic heterogenetity. *Hum. Genet.* **73**, 327-332.

Michaud, J., Brody, L.C., Steel, G., Fontaine, G., Martin, L.S., Valle, D. & Mitchell, G. (1992). Strand-separating conformational polymorphism analysis: efficacy of detection of point mutations in the human ornithine delta-aminotransferase gene. *Genomics* **13**, 389-394.

Mililli, M., Le Deist, F., de Saint Basile, G., Fischer, A., Fougereau, M. & Schiff, C. (1993). Bone marrow cells in X-linked agammaglobulinemia express pre-B-specific genes (lambda-like and V pre-B) and present immunoglobulin V-D-J gene usage strongly biased to a fetal-like repertoire. *J. Clin. Invest.* **91**, 1616-1629.

Miller, S.A., Dykes, D.D. & Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215

Mole, S.E., Iggo, R.D. & Lane, D.P. (1989). Using the polymerase chain reaction to modify expression plasmids for epitope mapping. *Nucleic Acids Res.* **17**, 3319

Monaco, A.P., Neve, R.L., Colletti-Feener, C., Bertelson, C.J., Kurnit, D.M. & Kunkel, L.M. (1986). Isolation of candidate cDNAs for portions of Duchenne muscular dystrophy gene. *Nature* **323**, 646-650.

Monaco, A.P., Walker, A.P., Millwood, I., Larin, Z. & Lehrach, H. (1992). A yeast artificial chromosome contig containing the complete Duchenne muscular dystrophy gene. *Genomics* **12**, 465-473.

Monafo, V., Maghnie, M., Terracciano, L., Valtorta, A., Massa, M. & Severi, F. (1991). X-linked agammaglobulinemia and isolated growth hormone deficiency. *Acta Paediatr. Scand.* **80**, 563-566.

Morgan, J.G., Dolganov, G.M., Robbins, S.E., Hinton, L.M. & Lovett, M. (1992). The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucleic Acids Res.* **20**, 5173-5179.

Moyret, C., Theillet, C., Puig, P.L., Moles, J.P., Thomas, G. & Hamelin, R. (1994). Relative efficiency of denaturing gradient gel electrophoresis and single strand conformation polymorphism in the detection of mutations in exons 5 to 8 of the p53 gene. *Oncogene* **9**, 1739-1743.

Moyzis, R.K., Torney, D.C., Meyne, J., Buckingham, J.M., Wu, J.R., Burks, C., Sirotkin, K.M. & Goad, W.B. (1989). The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* **4**, 273-289.

Murray, A.W., Schultes, N.P. & Szostak, J.W. (1986). Chromosome length controls mitotic chromosome segregation in yeast. *Cell* **45**, 529-536.

Musacchio, A., Gibson, T., Rice, P., Thompson, J. & Saraste, M. (1993). The PH domain: a common piece in the structural patchwork of signalling proteins. *TIBS* **18**, 343-348.

Myers, R.M., Lumelsky, N., Lerman, L.S. & Maniatis, T. (1985). Detection of single base substitutions in total genomic DNA. *Nature* **313**, 495-498.

Neil, D.L., Villasante, A., Fisher, R.B., Vetrie, D., Cox, B. & Tyler Smith, C. (1990). Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucleic Acids Res.* **18**, 1421-1428.

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H. & Caskey, C.T. (1989). *Alu* polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6686-6690.

Niemann Seyde, S., Slomski, R., Rininsland, F., Ellermeyer, U., Kwiatkowska, J. & Reiss, J. (1992). Molecular genetic analysis of 67 patients with Duchenne/Becker muscular dystrophy. *Hum. Genet.* **90**, 65-70.

Niwa, M., MacDonald, C.C. & Berget, S.M. (1992). Are vertebrate exons scanned during splice-site selection? *Nature* **360**, 277-280.

Noguchi, M., Yi, H., Rosenblatt, H.M., Filipovich, A.H., Adelstein, S., Modi, W.S., McBride, O.W. & Leonard, W.J. (1993). Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans. *Cell* **73**, 147-157.

Nolan, K.F., Schwaeble, W., Kaluz, S., Dierich, M.P. & Reid, K.B. (1991). Molecular cloning of the cDNA coding for properdin, a positive regulator of the alternative pathway of human complement. *Eur. J. Immunol.* **21**, 771-776.

Nolan, K.F., Kaluz, S., Higgins, J.M., Goundis, D. & Reid, K.B. (1992). Characterization of the human properdin gene. *Biochem. J.* **287**, 291-297.

O'Brien, M.C., Fukui, Y. & Hanafusa, H. (1990). Activation of the proto-oncogene p60c-src by point mutations in the SH2 domain. *Mol. Cell. Biol.* **10**, 2855-2862.

O'Reilly, M.A., Alterman, L.A., Malcolm, S., Levinsky, R.J. & Kinnon, C. (1992). Identification of CpG islands around the DXS178 locus in the region of the X-linked agammaglobulinaemia gene locus in Xq22. *Hum. Genet.* **90**, 275-278.

O'Reilly, M.A., Alterman, L.A., Zijlstra, J., Malcolm, S., Levinsky, R.J. & Kinnon, C. (1993a). Pulsed-field gel electrophoresis and radiation hybrid mapping analyses enable the ordering of eleven DNA loci in Xq22. *Genomics* **15**, 275-282.

O'Reilly, M.A., Sweatman, A.K., Bradley, L.D., Alterman, L.A., Lovering, R., Malcolm, S., Levinsky, R.J. & Kinnon, C. (1993b). Isolation and mapping of discrete

DXS101 loci in Xq22 near the X-linked agammaglobulinaemia gene locus. *Hum. Genet.* **91**, 605-608.

O'Reilly, M.A. & Kinnon, C. (1990). The technique of pulsed field gel electrophoresis and its impact on molecular immunology. *J. Immunol. Methods* **131**, 1-13.

Ohta, Y., Haire, R.N., Litman, R.T., Fu, S.M., Nelson, R.P., Kratz, J., Kornfeld, S.J., de la Morena, M., Good, R.A. & Litman, G.W. (1994). Genomic organisation and structure of Bruton agammaglobulinemia tyrosine kinase: Localisation of mutations associated with varied clinical presentations and course in X-chromosome linked agammaglobulinemia. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9062-9066.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. (1989a). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2766-2770.

Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. (1989b). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874-879.

Oto, M., Miyake, S. & Yuasa, Y. (1993). Optimization of nonradioisotopic single strand conformation polymorphism analysis with a conventional minislab gel electrophoresis apparatus. *Anal. Biochem.* **213**, 19-22.

Owerbach, D., Rutter, W.J., Martial, J.A., Baxter, J.D. & Shows, T.B. (1980). Genes for growth hormone, chorionic somatomammotropin and growth hormone-like gene on chromosome 17 in humans. *Science* **209**, 289-292.

Pagdett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. & Sharp, P.A. (1986). Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55**, 1119-1150.

Palmieri, G., Romano, G., Casamassimi, A., D'Urso, M., Little, R.D., Abidi, F.E., Schlessinger, D., Lagerstrom, M., Malmgren, H., Steen Bondeson, M.L., Pettersson, U. & Landegren, U. (1993). 1.5-Mb YAC contig in Xq28 formatted with sequence-tagged sites and including a region unstable in the clones. *Genomics* **16**, 586-592.

Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D. & Weissman, S.M. (1991). cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9623-9627.

Parkar, M., Lovering, R., Levinsky, R.J. & Kinnon, C. (1994). Genetic mapping of two loci, DXS454 and DXS458, with respect to the X-linked agammaglobulinemia gene locus. *Hum. Genet.* **93**, 89-90.

Parolini, O., Hejtmancik, J.F., Allen, R.C., Belmont, J.W., Lassiter, G.L., Henry, M.J., Barker, D.F. & Conley, M.E. (1993). Linkage analysis and physical mapping near

the gene for X-linked agammaglobulinemia at Xq22. *Genomics* **15**, 342-349.

Patanjali, S.R., Parimoo, S. & Weissman, S.M. (1991). Construction of a uniform-abundance (normalized) cDNA library [published erratum appears in Proc Natl Acad Sci U S A 1991 Jun 1; 88(11):5066]. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 1943-1947.

Pearl, E.R., Vogler, L.B., Okos, A.J., Crist, W.M., Lawton, A.R. & Cooper, M.D. (1978). B lymphocyte precursors in human bone marrow: an analysis of normal individuals and patients with antibody-deficiency status. *J. Immunol.* **120**, 1169-1175.

Pleiman, C.M., D'Ambrosio, D. & Cambier, J.C. (1994). The B cell antigen receptor complex: structure and signal transduction. *Immunol. Today* **15**, 393-399.

Prior, T.W., Papp, A.C., Snyder, P.J., Sedra, M.S., Western, L.M., Bartolo, C., Moxley, R.T. & Mendell, J.R. (1994). Heteroduplex analysis of the dystrophin gene: application to point mutation and carrier detection. *Am. J. Med. Genet.* **50**, 68-73.

Puck, J.M., Deschenes, S.M., Porter, J.C., Dutra, A.S., Brown, C.J., Willard, H.F. & Henthorn, P.S. (1993). The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-linked severe combined immunodeficiency, SCIDX1. *Hum. Mol. Genet.* **2**, 1099-1104.

Rastan, S. (1994). X chromosome inactivation and the Xist gene. *Curr. Opin. Genet. Dev.* **4**, 292-297.

Rawlings, D.J., Saffran, D.C., Tsukada, S., Largaespada, D.A., Grimaldi, J.C., Cohen, L., Mohr, R.N., Bazan, J.F., Howard, M., Copeland, N.G., Jenkins, N.A. & Witte, O.N. (1993). Mutation of unique region of Bruton's tyrosine kinase in immunodeficient XID mice. *Science* **261**, 358-361.

Ray, P.N., Belfall, B., Duff, C., Logan, C., Kean, V., Thompson, M.W., Sylvester, J.E., Gorski, J.L., Schmickel, R.D. & Worton, R.G. (1985). Cloning of the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy. *Nature* **318**, 672-675.

Reed, R. & Maniatis, T. (1986). A role for exon sequences and splice site proximity in splice site selection. *Cell* **46**, 681-690.

Ren, R., Mayer, B.J., Cicchetti, P. & Baltimore, D. (1993). Identification of a ten-amino acid proline-rich SH3 binding site. *Science* **259**, 1157-1161.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C. & Markham, A.F. (1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**, 2887-2890.

Robberson, B.L., Cote, G.J. & Berget, S.M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**, 84-94.

Roberts, R.G., Bobrow, M. & Bentley, D.R. (1992). Point mutations in the dystrophin gene. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2331-2335.

Roos,D. (1994). The genetic basis of chronic granulomatous disease. *Immunol. Rev.* **138**, 121-157.

Rommens, J.M., Iannuzzi, M.C., Kerem, B., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J.R., Tsui, L.-C. & Collins, F.S. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**, 1059-1065.

Rossiter, J.P., Young, M., Kimberland, M.L., Hutter, P., Ketterling, R.P., Gitschier, J., Horst, J., Morris, M.A., Schaid, D.J., de Moerloose, P., Sommer, S.S., Kazazian, H.H. & Antonarakis, S.E. (1994). Factor VIII gene inversions causing severe hemophilia A originate almost exclusively in male germ cells. *Hum. Mol. Genet.* **3**, 1035-1039.

Royer-Pokora, B., Kunkel, L.M., Monaco, A.P., Goff, S.C., Newberger, P.E., Baehner, R.L., Cole, F.S., Curnutte, J.T. & Orkin, S.H. (1986). Cloning the gene for an inherited human disorder - chronic granulomatous disease - on the basis of its chromosomal location. *Nature* **322**, 32-38.

Russell, D.W., Yamamoto, T., Schneider, W.J., Slaughter, C.J., Brown, M.S. & Goldstein, J.L. (1983). cDNA cloning of the bovine low density lipoprotein receptor: Feedback regulation of a receptor mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 7501-7505.

Saffran, D.C., Parolini, O., Fitch Hilgenberg, M.E., Rawlings, D.J., Afar, D.E., Witte, O.N. & Conley, M.E. (1994). Brief report: a point mutation in the SH2 domain of Bruton's tyrosine kinase in atypical X-linked agammaglobulinemia. *N. Engl. J. Med.* **330**, 1488-1491.

Sakai, N., Inui, K., Fujii, N., Fukushima, H., Nishimoto, J., Yanagihara, I., Isegawa, Y., Iwamatsu, A. & Okada, S. (1994). Krabbe disease: isolation and characterization of a full-length cDNA for human galactocerebrosidase. *Biochem. Biophys. Res. Commun.* **198**, 485-491.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular cloning: A laboratory manual.* Cold Spring Harbour Laboratory Press.

Saouaf, S.J., Mahajan, S., Rowley, R.B., Kut, S.A., Fargnoli, J., Burkhardt, A.L., Tsukada, S., Witte, O.N. & Bolen, J.B. (1994). Temporal differences in the activation of three classes of non-transmembrane protein tyrosine kinases following B cell antigen receptor surface engagement. *Proc. Natl. Acad. Sci. USA* **91**, 9524-9528.

Sarkar, G., Yoon, H.S. & Sommer, S.S. (1992a). Dideoxy fingerprinting (ddF): a rapid and efficient screen for the presence of mutations. *Genomics* **13**, 441-443.

Sarkar, G., Yoon, H.S. & Sommer, S.S. (1992b). Screening for mutations by RNA single-strand conformation polymorphism (rSSCP): comparison with DNA-SSCP. *Nucleic Acids Res.* **20**, 871-878.

Schinzel, A. (1984). *Catalogue of Unbalanced Chromosome Aberrations in Man.* Walter de Gryter, Berlin.

Scholz, R.B., Milde Langosch, K., Jung, R., Schlechte, H., Kabisch, H., Wagener, C. & Loning, T. (1993). Rapid screening for Tp53 mutations by temperature gradient gel electrophoresis: a comparison with SSCP analysis. *Hum. Mol. Genet.* **2**, 2155-2158.

Schwartz, D.C. & Cantor, C.R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67-75.

Serra, A. & Neri, G. (1990). Trisomy 21: conference report and 1990 update. *Am. J. Med. Genet.* **Supplement 7**, 11-19.

Shapiro, M.B. & Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155-7174.

Sheffield, V.C., Cox, D.R., Lerman, L.S. & Myers, R.M. (1989). Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 232-236.

Sheffield, V.C., Beck, J.S., Kwitek, A.E., Sandstrom, D.W. & Stone, E.M. (1993). The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* **16**, 325-332.

Shimizu, T., Toguchida, J., Kato, M.V., Kaneko, A., Ishizaki, K. & Sasaki, M.S. (1994). Detection of mutations of the RB1 gene in retinoblastoma patients by using exon-by-exon PCR-SSCP analysis. *Am. J. Hum. Genet.* **54**, 793-800.

Sideras, P., Muller, S., Shiels, H., Jin, H., Khan, W.N., Nilsson, L., Parkinson, E., Thomas, J.D., Branden, L., Larsson, I., Paul, W.E., Rosen, F.S., Alt, F.W., Vetrie, D., Smith, C.I.E. & Xanthopoulos, K.G. (1994). Genomic organisation of mouse and human Bruton's agammaglobulinemia tyrosine kinase (*Btk*) locus. *J. Immunol.* (In Press).

Siliciano, J.D., Morrow, T.A. & Desiderio, S.V. (1992). itk, a T-cell-specific tyrosine kinase gene inducible by interleukin 2. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 11194-11198.

Silverman, G.A., Ye, R.D., Pollock, K.M., Sadler, J.E. & Korsmeyer, S.J. (1989). Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7485-7489.

Sitz, K.V., Burks, A.W., Williams, L.W., Kemp, S.F. & Steele, R.W. (1990). Confirmation of X-linked hypogammaglobulinemia with isolated growth hormone deficiency as a disease entity. *J. Pediatr.* **116**, 292-294.

Skare, J., Wu, B.-L., Madan, S., Pulijaal, V., Purtilo, D., Haber, D., Nelson, D., Sylla, B., Grierson, H., Nitowsky, H., Glaser, J., Wissink, J., White, B., Holden, J., Housman, D., Lenoir, G., Wyandt, H. & Milunsky, A. (1993). Characterisation of three overlapping deletions causing X-linked lymphoproliferative disease. *Genomics* **16**, 254-255.

Sleister, H.M., Mills, K.A., Blackwell, S.E., Killary, A.M., Murray, J.C. & Malone, R.E. (1992). Construction of a human chromosome 4 YAC pool and analysis of artificial chromosome stability. *Nucleic Acids Res.* **20**, 3419-3425.

Smale, S.T. & Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell* **57**, 103-113.

Smith, C.I.E., Baskin, B., Humire-Greiff, P., Zhou, J., Olsson, P.G., Maniar, H.S., Kjellen, P., Lambris, J.D., Christensson, B., Hammarstrom, L., Bentley, D., Vetrie, D., Islam, K.B., Vorechovsky, I. & Sideras, P. (1994a). Expression of Bruton's agammaglobulinemia tyrosine kinase gene, *BTK*, is selectively down-regulated in T lymphocytes and plasma cells. *J. Immunol.* **152**, 557-565.

Smith, C.I.E., Islam, K.B., Vorechovsky, I., Olerup, O., Wallin, E., Rabbani, H., Baskin, B. & Hammarstrom, L. (1994b). X-Linked Agammaglobulinemia and Other Immunoglobulin Deficiencies. *Immunological Reviews* **138**, 159-183.

Smith, D.B. & Johnson, K.S. (1988). Single step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-s-transferase. *Gene* **15**, 31-40.

Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J., Neel, B.G., Birge, R.B., Fajardo, J.E., Chou, M.M., Hanafusa, H., Schaffhausen, B. & Cantley, L.C. (1993). SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767-778.

Superti Furga, G., Fumagalli, S., Koegl, M., Courtneidge, S.A. & Draetta, G. (1993). Csk inhibition of c-Src activity requires both the SH2 and SH3 domains of Src. *EMBO J.* **12**, 2625-2634.

Sweatman, A.K. (1994). Physical Mapping Studies of the X-linked Agammaglobulinemia Locus in Xq22. PhD thesis, Univerity of London.

Sweatman, A.K., Bradley, L.A.D., Lovering, R.C., O'Reilly, M.A., Levinsky, R.J. &

Kinnon, C. (1994). Physical mapping in the region of the Bruton's tyrosine kinase and alpha-galactosidase A gene loci in proximal Xq22. *Hum. Genet.* (in press).

Takahashi, N. & Ko, M.S.H. (1994). Toward a whole cDNA catalog: construction of an equalized cDNA library from mouse embryos. *Genomics* **23**, 202-210.

Takeshita, T., Asao, H., Ohtani, K., Ishii, N., Kumaki, S., Tanaka, N., Munakata, H., Nakamura, M. & Sugamura, K. (1992). Cloning of the gamma chain of the human IL-2 receptor. *Science* **257**, 379-382.

Taniguchi, T. & Minami, Y. (1993). The IL2/IL2 receptor system: a current overview. *Cell* **73**, 5-8.

Tarlinton, D. (1994). B-cell differentiation in the bone marrow and the periphery. *Immunol. Rev.* **137**, 203-229.

Tazi, J. & Bird, A. (1990). Alternative chromatin structure at CpG islands. *Cell* **60**, 909-920.

Tedder, T.F., Crain, M.J., Kubagawa, H., Clement, L.T. & Cooper, M.D. (1985). Evaluation of lymphocyte differentiation in primary and secondary immunodeficiency diseases. *J. Immunol.* **135**, 1786-1791.

Thomas, J.D., Sideras, P., Smith, C.I., Vorechovsky, I., Chapman, V. & Paul, W.E. (1993). Colocalization of X-linked agammaglobulinemia and X-linked immunodeficiency genes. *Science* **261**, 355-358.

Timmers, E., de Weers, M., Alt, F.W., Hendriks, R.W. & Schuurman, R.K. (1991). X-linked agammaglobulinemia. *Clin. Immunol. Immunopathol.* **61**, S83-S93.

Touhara, K., Inglese, J., Pitcher, J.A., Shaw, G. & Lefkowitz, R.J. (1994). Binding of G protein beta gamma-subunits to pleckstrin homology domains. *J. Biol. Chem.* **269**, 10217-10220.

Trask, B.J. (1991). Fluoresence *in situ* hybridisation: applications in cytogenetics and gene mapping. *Trends Genet.* **7**, 149-155.

Tribioli, C., Mancini, M., Plassart, E., Bione, S., Rivella, S., Sala, C., Torri, G. & Toniolo, D. (1994). Isolation of new genes in distal Xq28: transcriptional map and identification of a human homologue of the ARD1 N-acetyl transferase of *Saccharomyces cerevisiae*. *Hum. Mol. Genet.* **3**, 1061-1067.

Tsukada, S., Saffran, D.C., Rawlings, D.J., Parolini, O., Allen, R.C., Klisak, I., Sparkes, R.S., Kubagawa, H., Mohandas, T., Quan, S., Belmont, J.W., Cooper, M.D., Conley, M.E. & Witte, O.N. (1993). Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell* **72**, 279-290.

Umetsu, D.T., Ambrosino, D.M., Quinti, I., Siber, G.R. & Geha, R.S. (1985). Recurrent sinopulmonary infection and impaired antibody response to bacterial capsular polysaccharide antigen in children with selective IgG-subclass deficiency. *N. Engl. J. Med.* **313**, 1247-1251.

Urlaub, G., Mitchell, P.J., Ciudad, C.J. & Chasin, L.A. (1989). Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. *Mol. Cell. Biol.* **9**, 2868-2880.

Vetrie, D., Bentley, D., Bobrow, M. & Harris, A. (1993a). Physical mapping shows close linkage between the alpha-galactosidase A gene (GLA) and the DXS178 locus. *Hum. Genet.* **92**, 95-99.

Vetrie, D., Bobrow, M. & Harris, A. (1993b). Construction of a 5.2-megabase physical map of the human X chromosome at Xq22 using pulsed-field gel electrophoresis and yeast artificial chromosomes. *Genomics* **15**, 631-642.

Vetrie, D., Vorechovsky, I., Sideras, P., Holland, J., Davies, A., Flinter, F., Hammarstrom, L., Kinnon, C., Levinsky, R., Bobrow, M., Smith, C.I.E. & Bentley, D. (1993c). The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. *Nature* **361**, 226-233.

Vetrie, D., Kendall, E., Coffey, A., Hassock, S., Collins, J., Todd, C., Lehrach, H., Bobrow, M., Bentley, D.R. & Harris, A. (1994). A 6.5-Mb yeast artificial chromosome contig incorporating 33 DNA markers on the human X chromosome at Xq22. *Genomics* **19**, 42-47.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Preisinger, A.C., Willard, H.F., Michelson, A.M., Riggs, A.D. & Orkin, S.H. (1987). Clonal analysis using recombinant DNA probes from the X-chromosome. *Cancer Res.* **47**, 4806-4813.

Vorechovsky, I., Zhou, J.-N., Vetrie, D., Bentley, D., Bjorkander, J., Hammerstrom, L. & Smith, C.I.E. (1993). Molecular diagnosis of X-linked agammaglobulinemia. *Lancet* **341**, 1153.

Vorechovsky, I. (1994). SSCP at the Btk locus. *Hum. Mol. Genet.* **3**, 1031.

Vorechovsky, I., Vetrie, D., Holland, J., Bentley, D.R., Thomas, K., Zhou, J.-N., Notarangelo, L.D., Plebani, A., Fontan, G., Ochs, H.D., Hammarstrom, L., Sideras, P. & Smith, C.I.E. (1994). Isolation of cosmid and cDNA clones in the region surrounding the Btk gene at Xq21.3-q22. *Genomics* **21**, 517-524.

Wada, M., Little, R.D., Abidi, F., Porta, G., Labella, T., Cooper, T., Della Valle, G., D'Urso, M. & Schlessinger, D. (1990). Human Xq24-Xq28: approaches to mapping with yeast artificial chromosomes. *Am. J. Hum. Genet.* **46**, 95-106.

Wadelius, C., Pigg, M., Sundvall, M., Sjoholm, A.G., Goonewardena, P., Kuijper, E.J., Tijssen, C.C., Jansz, A., Spath, P.J., Schaad, U.B., Tranebaerg, L., Nielsen,

H.E., Soderstrom, C., Anneren, G. and Petersson, U. (1992). Linkage analysis in properdin deficiency families: refined location in proximal Xp. *Clin. Genet.* **42**, 8-12.

Waksman, G., Kominos, D., Robertson, S.C., Pant, N., Baltimore, D., Birge, R.B., Cowburn, D., Hanafusa, H., Mayer, B.J., Overduin, M., Resh, M.D., Rios, C.B, Silverman, L. and Kuriyan, J. (1992). Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **358**, 646-653.

Waksman, G., Shoelson, S.E., Pant, N., Cowburn, D. & Kuriyan, J. (1993). Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell* **72**, 779-790.

Wallace, M.R., Marchuk, D.A., Andersen, L.B., Letcher, R., Odeh, H.M., Saulino, A.M., Fountain, J.W., Brereton, A., Nicholson, J., Mitchell, A.L. (1990). Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients [published erratum appears in Science 1990 Dec 21;250(4988):1749]. *Science* **249**, 181-186.

Weber, F., de Villiers, J. & Schaffner, W. (1984). An SV40 "enhancer trap" incorporates exogenous enhancers from its own sequences. *Cell* **36**, 983-992.

Weber, J.L., Kwitek, A.E., May, P.E., Polymeropoulos, M.H. & Ledbetter, S. (1990). Dinucleotide repeat polymorphisms at the DXS453, DXS454 and DXS458 loci. *Nucleic Acids Res.* **18**, 4037.

Weeda, G., Hoeijmakers, J.H.J. & Bootsma, D. (1993). Genes controlling nucleotide excision repair in eukaryotic cells. *Bioessays* **15**, 249-258.

White, M.B., Carvalho, M., Derse, D., O'Brien, S.J. & Dean, M. (1992). Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* **12**, 301-306.

Yao, L., Kawakami, Y. & Kawakami, T. (1994). The pleckstrin homolgy domain of Bruton tyrosine kinase interacts with protein kinase C. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9175-9179.

Yoon, H.S., Hajduk, P.J., Petros, A.M., Olejniczak, E.T., Meadows, R.P. & Fesik, S.W. (1994). Solution structure of a pleckstrin-homology domain. *Nature* **369**, 672-675.

Yu, H., Rosen, M.K., Shin, T.B., Seidel Dugan, C., Brugge, J.S. & Schreiber, S.L. (1992). Solution structure of the SH3 domain of Src and identification of its ligand-binding site. *Science* **258**, 1665-1668.

Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W. & Schreiber, S.L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933-945.

Zhu, D., Mitchell, T.N. & Maumenee, I.H. (1991). Mapping of the X-linked FG syndrome to Xq21.31-Xq22. *Cytogenet. Cell Genet.* **58**, A27533 (Abstract).

Zhu, Q., Zhang, M., Rawlings, D.J., Vihinen, M., Hagemann, T., Saffran, D.C., Kwan, S.-P., Nilsson, L., Smith, C.I.E., Witte, O.N., Chen, S.-H. & Ochs, H.D. (1994a). Deletion within the Src homology domain 3 of Bruton's tyrosine kinase resulting in X-linked agammaglobulinemia. *J. Exp. Med.* **180**, 461-470.

Zhu, Q., Zhang, M., Winkelstein, J., Chen, S.-H. & Ochs, H.D. (1994b). Unique mutations of Bruton's tyrosine kinase in fourteen unrelated X-linked agammaglobulinemia families. *Hum. Mol. Genet.* **3**, 1899-1900.

Zonana, J., Roberts, S.H., Thomas, N.S. & Harper, P.S. (1988). Recognition and reanalysis of a cell line from a manifesting female with X linked hypohidrotic ectodermal dysplasia and an X; autosome balanced translocation. *J. Med. Genet.* **25**, 383-386.

# APPENDICES

## Appendix I DNA probes used in this study

| Locus | Probe | Washing Conditions | Reference |
|---|---|---|---|
| DXS3 | p19-2 | 1 x SSC | Bruns *et al.* 1982 |
| DXS366 | RX329 | 0.5 x SSC | Barker *et al.* 1991 |
| DXS442 | RX276 | 1 x SSC | Barker *et al.* 1991 |
| GLA | GLx6 | 0.5 x SSC | This study |
| 5D8 | 5D8 | 0.5 x SSC | Vetrie *et al.* 1993 |
| *BTK* | 7C11 | 0.5 x SSC | Vetrie *et al.* 1993 |
| *BTK* | 6G11 | 0.5 x SSC | Vetrie *et al.* 1993 |
| DXS178 | p212XT | 0.5 x SSC | Lovering *et al.* 1993a |
| DXS178 | 212PCR | 1 x SSC | This study |
| DXS265 | pKZ033H4 | 1 x SSC | Dietz Band *et al.* 1990 |
| DXS101 | GPS | 1 x SSC | O'Reilly *et al.* 1993 |
| DXS101 | B1.4 | 1 x SSC | O'Reilly *et al.* 1993 |
| DXS101 | cX52.5 | 3 x SSC | Hofker *et al.* 1987 |

## Appendix II Clinical details of patients

| Patient | Diagnosis | B cells | IgG | IgA | IgM | Family | Comments |
|---|---|---|---|---|---|---|---|
| BB | XLA | <1% | 0.21<br>(5.4-16.1) | undetectable | 0.36<br>(0.5-1.8) | no history | |
| SB | XLA | <1% | n/a | n/a | n/a | 3 brothers/2 affected | |
| LT | XLA | n/a | n/a | n/a | n/a | | |
| JG | XLA | <1% | 11.98* | 0.04<br>(0.8-2.8) | <0.07<br>(0.5-1.9) | 4 brothers/3 affected | |
| CAB | XLA | n/a | n/a | n/a | n/a | large pedigree<br>affected cousin | |
| FGV | XLA | <1% | n/a | n/a | n/a | | |
| NAD | XLA | <1% | 7.36* | 0.15<br>(0.3-1.2) | <0.08<br>(0.5-2.2) | 1 affected brother | |
| 276 | XLA | n/a | n/a | n/a | n/a | Mensink *et al.* 1984<br>de Weers *et al.* 1994 | |
| AC | XLA | n/a | n/a | n/a | n/a | maternal-side<br>uncle died young | |
| JP | XLA | <1% | <1.0 | <0.2 | <0.1 | sporadic | |
| FM | XLA | n/a | n/a | n/a | n/a | 2 brothers affected | |
| TF | XLA | 1% | 8.09<br>(4.9-16.1) | 0.46<br>(0.4-2.0) | 0.25<br>(0.5-2.0) | brother of JF and KF | IgG1 488 (280-900)<br>IgG2 <3 (59-251)<br>IgG3 30 (38-78)<br>IgG4 12 (7.5-26) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| JF | XLA | <1% | 3.84<br>(3.7-15.8) | 0.16<br>(0.3-1.3) | 0.22<br>(0.5-2.2) | brother of TF and KF | IgG1 204 (315-759)<br>IgG2 <3 (61-255)<br>IgG3 <2 (23-80)<br>IgG4 <2 (6-26) |
| KF | XLA | 2% | 4.00<br>(2.1-7.7) | <0.03<br>(0.05-0.4) | 0.07<br>(0.15-0.7) | brother of TF and JF | |
| AJ | XLA | <1% | n/a | n/a | n/a | sporadic | |
| CHD | XLA | <1% | n/a | n/a | n/a | | |
| AP | XLA | <1% | 1.2<br>(63-160) | 1.0<br>(25-117) | 5.2<br>(50-220) | 3 brothers/3 affected<br>Mother non-random | U/ml |
| OS | IgG A deficiency | n/a normal | 6.76<br>(3.1-13.8) | 0.15<br>(0.3-1.2) | 0.98<br>(0.5-2.5) | two brothers died of infections/ 2 healthy sisters | no IgG1 or IgG2 response after immunisation to pneumoccocal polysaccharide |
| RF | IgA and IgG2 & 4 subclass deficiency | n/a normal | 4.73<br>(3-10.9) | 0.17<br>(0.2-0.7) | 0.54<br>(0.6-2.1) | brother died of infection /<br>2 sisters healthy | no IgG1 or IgG2 response after immunisation to pneumococcal polysaccharide |
| DW | Hypoγ | 5% | n/a | 0.06<br>(0.3-1.2) | 0.08<br>(0.5-2.2) | | |

Normal ranges for immunoglobulin measurements shown in brackets. * indicates patient receiving gammaglobulin therapy when measurement taken. XLA diagnoses made on the basis of a lack of B cells and normal cell mediated immunity. All immunoglobulin measurements in g/L unless otherwise stated, except of IgG subclass measurements which are in mg/dl. n/a indicates data not available.

## Appendix III Summary of restriction fragments from YAC 178-2

(a) *Bss*HII partial restriction fragments

| Fragment size (kb) | YAC R end | DXS 366 | DXS 442 | 5D8 | *BTK* 6G11 | *GLA* | YAC L end |
|---|---|---|---|---|---|---|---|
| 565 | + | + | + | - | - | - | - |
| 515 | - | + | + | - | - | - | - |
| 490 | - | + | + | - | - | - | - |
| 480 | - | o | + | o | + | + | + |
| 420 | - | + | + | - | - | - | - |
| 410 | + | - | - | - | - | - | - |
| 375 | - | + | - | - | - | - | - |
| 370 | - | - | + | + | + | + | + |
| 325 | - | - | - | + | + | + | + |
| 250 | - | + | + | - | - | - | - |
| 245 | - | - | - | + | + | + | - |
| 220 | - | - | - | + | + | + | - |
| 195 | - | - | - | + | + | + | + |
| 165 | - | - | - | + | + | + | + |
| 155 | - | + | + | - | - | - | - |
| 155 | - | - | - | + | + | + | + |
| 145 | + | - | - | - | - | - | - |
| 140 | - | - | + | - | - | - | - |
| 110 | - | + | - | - | - | - | - |
| 90 | - | - | - | + | + | + | - |
| 80 | - | - | - | + | + | - | - |
| 75 | - | - | - | - | - | + | + |
| 65 | - | - | - | - | - | (+) | (+) |
| 45 | - | - | + | - | - | - | - |
| 40 | - | - | - | + | + | + | - |
| 30 | - | - | - | + | + | - | - |
| 10 | (+) | - | - | - | - | (+) | (+) |

*(b) SfiI* partial restriction fragments

| Fragment size (kb) | YAC R end | DXS 366 | DXS 442 | 5D8 | *BTK* 6G11 | *GLA* | YAC L end |
|---|---|---|---|---|---|---|---|
| nd | - | - | +* | - | - | - | - |
| nd | - | - | +* | - | +* | - | - |
| 405 | (+) | - | (+ | o | + | o) | - |
| 355 | - | + | + | - | - | - | - |
| 345 | + | - | - | - | - | - | - |
| 315 | - | - | - | o | + | o | - |
| 300 | - | - | + | - | - | - | - |
| 280 | - | + | + | - | - | - | - |
| 260 | - | - | - | + | + | + | + |
| 235 | - | - | - | + | + | + | + |
| 210 | + | - | - | - | - | - | - |
| 205 | - | - | - | + | + | + | + |
| 180 | + | - | - | - | - | - | - |
| 160 | - | + | + | - | - | - | - |
| 155 | - | - | - | + | + | + | + |
| 130 | - | - | + | - | - | - | - |
| 110 | + | - | - | - | - | - | - |
| 105 | - | - | - | + | + | + | - |
| 50 | (+) | - | - | - | - | - | (+) |
| 30 | - | + | - | - | - | - | - |
| 10 | - | - | - | - | - | - | + |

+ indicates a fragment hybridising to the probe specific for the indicated locus on Southern blot analysis and - indicates no hybridisation. Brackets indicate common fragments when coincidental fragment sizes are suspected. * indicates that the data was incomplete due to faint hybridisation signals and that these data were not included in the proposed restriction digest map. o indicates that the hybridisation signal was very faint but the band was taken to be present on the basis of other data. nd indicates not determined.

# PUBLICATIONS

Data presented in this thesis has been published in the following papers:

Lovering, R., Middleton, H.R., O'Reilly, M.-A.J., Genet, S., Parkar, M., Sweatman, A.K., **Bradley, L.D.**, Alterman, L.A., Malcolm, S., Morgan, G., Levinsky, R. and Kinnon, C. (1993) Genetic linkage analysis identifies new proximal and distal flanking markers for the X-linked agammaglobulinemia gene locus, refining its localisation in Xq22. *Hum. Mol. Genet.* **2**:139-141

O'Reilly, M.-A.J., Sweatman, A.K., **Bradley, L.D.**,Alterman, L.A., Lovering, R.,Malcolm, S., Levinsky, R. and Kinnon, C. (1993) Isolation and mapping of discrete DXS101 loci in Xq22 in Xq22 near the X-linked agammaglobulinemia gene locus. *Hum. Genet.* **91**:605-608

**Bradley, L.A.D.**, Sweatman, A.S., Lovering, R.C., Jones, A.M., Morgan, G., Levinsky, R.J. and Kinnon, C. (1994) Mutation Detection in the X-linked agammaglobulinemia gene, Btk, using single strand conformation polymorphism analysis. *Hum. Mol. Genet.* **3**:79-83

Katz, F.E., Lovering, R.C., **Bradley, L.A.D.**, Rigley, K.P., Brown, D., Cotter, F.C., Chessells, J.M., Levinsky, R.J. and Kinnon, C. (1994) Expression of the X-linked agammaglobulinemia gene, Btk, in B cell acute lymphoblastic leukemia. *Leukemia* **8**:574-577

Sweatman, A.S., **Bradley, L.A.D.**, Lovering, R.C., O'Reilly, M.-A.J., Levinsky, R. and Kinnon, C. A physical map in the region of the X-linked agammaglobulinemia and $\alpha$ galactosidase loci in proximal Xq22. *Hum. Genet.* (in press).

Human Molecular Genetics, Vol. 2, No. 2 139–141

# Genetic linkage analysis identifies new proximal and distal flanking markers for the X-linked agammaglobulinemia gene locus, refining its localization in Xq22

Ruth Lovering, Helen R.Middleton-Price[2], Marie-Anne J.O'Reilly, Sally A.Genet[2], Mohammed Parkar, Angela K.Sweatman, Linda D.Bradley, Lesley A.Alterman, Sue Malcolm[1], Gareth Morgan, Roland J.Levinsky and Christine Kinnon*

Molecular Immunology and [1]Molecular Genetics, Division of Cell and Molecular Biology and [2]Clinical Genetics, Division of Public Health, Institute of Child Health, University of London, 30 Guilford Street, London WC1N 1EH, UK

Received October 13, 1992; Revised and Accepted December 3, 1992

## ABSTRACT

**Genetic linkage analysis has been instrumental in mapping the gene for X-linked agammaglobulinemia (XLA) to the proximal long arm of the human X chromosome, to Xq22. Due to the relative rarity of this disease the localization of the gene within Xq22 has remained imprecise. We have investigated twenty-nine families affected by XLA and have found no recombinants with the DXS178 locus in over 30 informative meioses. DXS178 is now the most reliable and informative locus for use in pre-natal diagnosis and carrier detection of XLA. In addition, we have identified new closely linked proximal and distal flanking markers for XLA, DXS442 and DXS101, respectively. These loci are separated by 2cM, considerably reducing the extent of DNA within which the XLA locus can be contained. This will open up the way for more directed positional cloning efforts for the isolation of the XLA gene.**

## INTRODUCTION

X-linked agammaglobulinemia (XLA) is a rare immunodeficiency disorder, whereby affected boys have no circulating antibodies and thus, no humoral immunity. Since the nature of the protein product of the XLA locus is not known, attempts are being made to identify the gene by positional cloning. No XLA patients have been identified who have cytogenetic abnormalities or who are deleted for any of the closely linked markers, thus genetic linkage analysis has proved to be the most useful first step in this process.

Genetic linkage analysis of affected pedigrees has revealed an overall consensus order of loci in Xq22 as: cen-DXS3-(XLA, DXS178)-DXS94-DXS17-tel, where there are no recombinations between the disease and the DXS178 locus in over 30 informative meioses to date[1,2]. This map has recently been extended to include additional markers. A preliminary study showed no recombinations between DXS265, DXS327, DXS366, DXS442 and the XLA locus in the nine informative families studied[3]. Since DXS265 lies within 5kb of DXS178, and is in linkage disequilibrium with this locus[4], it would not be expected to show recombination with the disease. The DXS101 locus lies within 1Mb of DXS178 but has not so far been placed proximal or distal relative to DXS178 and has been demonstrated not to have recombined with XLA in ten informative meioses[5,6].

The localization of these additional loci in Xq22 was determined in a genetic linkage study of two Alport syndrome families, families known to be defective in the COL4A5 gene which lies distal to DXS17[7]. This study gave the following order of loci: cen-DXS3-DXS366-DXS442-DXS101-DXS17-DXS358-tel, with DXS94, DXS178, DXS265 and DXS327 unplaced on this map but lying between DXS442 and DXS358[7]. Physical mapping of the breakpoints in hybrid cell lines assigned the DXS94, DXS178, DXS265 and DXS327 loci to the same interval as DXS17, DXS101, DXS366 and DXS442, placing them all between DXS3 and COL4A5[7].

We have now studied a total of twenty-nine XLA families, consisting of 477 individuals, using these markers. This is the largest genetic linkage study of XLA families to date and uses three new markers which have not previously been employed in multipoint analysis with XLA (DXS101, DXS366 and DXS442). Data obtained from this study have allowed us to construct a refined genetic map of this region: cen-DXS3-DXS366-DXS442-(DXS178, DXS265, XLA)-DXS101-DXS94-DXS17-tel, with the DXS327 locus mapping distal to DXS178. This genetic map identifies two new flanking markers for the XLA locus: DXS442 as the closest proximal flanking marker and DXS101 as the closest distal flanking marker. These two markers are separated by a genetic distance of 2cM[7].

## RESULTS AND DISCUSSION

### Two point LOD scores

Two point lod scores between XLA and eight loci are presented in Table 1. Marker-marker two point lod scores have been submitted to GDB. DXS265 was not included in this study since it lies within 5kb of DXS178 and is in almost complete linkage disequilibrium with this locus[4]. The marker-marker and marker-disease recombination events are shown schematically in Fig. 1. Eleven recombinant chromosomes were identified in this study and we could recognise the XLA status of the recombinant chromosome in four of these cases (chromosomes 3, 7, 8, 11,

* To whom correspondence should be addressed

**Table 1.** Two-point lod scores between XLA and 8 loci in Xq22

| XLA | DXS3 | DXS366 | DXS442 | DXS178 | DXS101 | DXS94 | DXS17 | DXS327 |
|---|---|---|---|---|---|---|---|---|
| $Z_{max}$ | 8.90 | 8.32 | 6.19 | 10.48 | 9.07 | 9.50 | 10.96 | 9.11 |
| $\theta_{max}$ | 0.04 | 0.04 | 0.03 | 0.00 | 0.02 | 0.00 | 0.02 | 0.00 |
| $\theta$ | lod score | | | | | | | |
| 0.000 | −188.82 | −192.75 | −92.39 | 10.48 | −90.56 | 9.50 | −87.63 | 9.11 |
| 0.001 | 6.64 | 5.94 | 5.19 | 10.46 | 8.22 | 9.48 | 10.14 | 9.09 |
| 0.010 | 8.42 | 7.75 | 6.04 | 10.27 | 9.01 | 9.30 | 10.92 | 8.93 |
| 0.050 | 8.85 | 8.29 | 6.07 | 9.40 | 8.76 | 8.49 | 10.59 | 8.19 |
| 0.100 | 8.20 | 7.80 | 5.51 | 8.26 | 7.86 | 7.45 | 9.57 | 7.19 |
| 0.200 | 6.22 | 6.10 | 4.03 | 5.89 | 5.72 | 5.36 | 7.14 | 5.04 |
| 0.300 | 3.95 | 4.07 | 2.48 | 3.52 | 3.46 | 3.31 | 4.49 | 2.80 |
| 0.400 | 1.74 | 1.95 | 1.06 | 1.42 | 1.33 | 1.45 | 1.88 | 0.99 |

Maximum lod score ($Z_{max}$) at maximum likelihood estimate (MLE) of the recombination fraction ($\theta_{max}$).

Fig. 1). Two loci which had previously not been mapped with respect to XLA have shown recombination with the disease locus. A proximal recombination event with DXS366 and DXS442 is shown by chromosome 8, while a distal recombination is shown with DXS101 by chromosome 11.

DXS178 is the most accurate marker available for carrier detection and pre-natal diagnosis (PND) for XLA since cumulative data have narrowed the 95% confidence interval to within 3% recombination on two point analysis alone. With these and other data[1,2], the total cumulative lod score between XLA and DXS178 now stands at over 21 with no recombinations found in 70 meioses. The probe 212XT (at the DXS178 locus) is a useful marker for clinical application since it detects an RFLP with 44% heterozygosity which results in a high proportion of informative females. Furthermore, with the highly informative CA repeat polymorphisms recently described at this locus[17,18], the majority of women wishing for carrier assessment should now be informative for one or other of these polymorphisms. Should flanking loci be required for assessment of carrier status and PND in affected families then there are now six closely linked RFLPs available. Based on our data and additional data[1-3], the cumulative two point lod score between XLA and DXS3 is 15.06 ($\theta$=0.05), between XLA and DXS94 is 11.26 ($\theta$=0.05) and between XLA and DXS17 is 13.04 ($\theta$=0.05). Cumulative two point lod scores between XLA and new markers in Xq22 are as follows: XLA-DXS366, lod=12.57 (at $\theta$=0.05); XLA-DXS442 lod=7.55 (at $\theta$=0.05); XLA-DXS327 lod=11.62 (at $\theta$=0). This study supercedes those of Malcolm *et al.*[14] and Lau *et al.*[15]

### Marker locus order within Xq22

For the multipoint linkage analysis with XLA the order of loci in the Xq22 region had first to be fixed. From an extensive genetic and physical mapping study of loci near the X-linked Alport syndrome locus (COL4A5)[7] the order of some of the loci used in this study could be fixed as: cen-DXS3-3cM-DXS366-1cM-DXS442-2cM-DXS101-1cM-DXS17-tel. However, the loci DXS94, DXS178 and DXS327 were not placed in this study, other than being between DXS3 and COL4A5[7]. The DXS94 locus was placed proximal to DXS17 on the basis of a single recombination event[9]. Our physical mapping data suggests that DXS94 is distal to DXS101 as the DXS17 and DXS94 loci lie on the same 2.7 Mb *Mlu*I fragment which does not contain the DXS101 locus[6,10]. Since we observed no recombinants between DXS94 and its flanking loci, DXS17 and DXS101, we placed

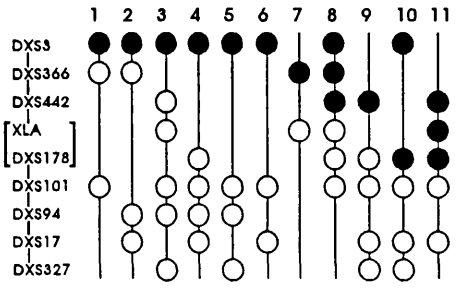



**Figure 1.** Representation of recombinant chromosomes in XLA families. Informative loci are represented by circles. Filled or empty circles correspond to loci derived from one or the other of the maternal X chromosomes.

it in the middle of this interval. The revised map: cen-DXS3-3cM-DXS366-1cM-DXS442-2cM-DXS101-0.5cM-DXS94-0.5cM-DXS17-tel was then used to estimate the position of the DXS178 locus using the LINKMAP section of LINKAGE[11]. DXS178 was mapped at the midpoint within the interval between the DXS442 and DXS101 loci with the highest multipoint lod score of 14.25 and a significantly greater likelihood ($>10^4$) than any other interval except distal to DXS17. Since other data from this region clearly places DXS178 proximal to DXS17[2,9], the loci order DXS442-1cM-DXS178-1cM-DXS101 was fixed for further analysis. Using the LINKMAP program the DXS327 locus was found to lie distal to DXS178 with the greatest probability (odds $>10^2$) and with a multipoint lod score of 15.5 at DXS101. However, since we observed no recombinations between this locus and loci distal to DXS178, in the chromosomes studied, its position could not be further refined within the map and so it was omitted from further analysis. The final order of loci for analysis with XLA was: cen-DXS3-3cM-DXS366-1cM-DXS442-1cM-DXS178-1cM-DXS101-0.5cM-DXS94-0.5cM-DXS17-tel.

### Multipoint linkage analysis with XLA

The position of XLA on the fixed order map relative to seven loci from Xq22 has been determined. The most likely position for XLA is at DXS178 with which it gives a multipoint lod score

of 21.6. The odds in favour of this position over all other positions are $>10^3$ except for its next most likely position 1cM distal to DXS17 where they are $>10^2$. However, the consensus genetic map rules out XLA being distal to DXS17[1,2,12]. Recombinations with DXS442 and DXS101 (chromosomes 8 and 11, Fig. 1) identify these loci as the closest proximal and distal flanking markers respectively. The order of loci in this region is now: cen-DXS3-DXS366-DXS442-(DXS178, XLA)-DXS101-DXS94-DXS17-tel.

The results of this study suggest that all further efforts to identify gene candidates for XLA should concentrate on the region between DXS442 and DXS101. Estimates for the genetic distances between loci in this region vary quite considerably. Although the genetic distance between the previously identified flanking markers, DXS3 and DXS94, was estimated to be 11.5cM[2], this was within an overall genetic map in which the distance between the DXS3 and DXS17 loci was 15cM. ILINK analysis of the data presented here suggests a distance of 12cM between DXS3 and DXS17. However, a much larger study by Barker et al[7] which maps several new loci within this interval estimates this distance to be only 7cM. It is within this map that the new flanking markers DXS442 and DXS101, separated by 2cM, are placed. So, while it is difficult to directly compare the genetic distance between the previous flanking markers for the XLA locus with the distance between the new flanking markers identified in this study, it is clear that the new flanking markers for XLA considerably reduce the genetic distance over which the XLA locus can lie.

The distance between the closest flanking markers for XLA, DXS442 and DXS101, define the confidence interval for XLA at DXS178. The genetic distance between DXS442 and DXS101 is estimated as being 2cM by a large study of two families with Alport syndrome[7]. While our ILINK analysis estimates this distance to be somewhat larger at 4cM, we have used 2cM in the multipoint analysis since it is derived from a larger body of data. As there appears to be some discrepancy in the estimates of the genetic distance between DXS442 and DXS101, we are using the 95% confidence interval of 3% recombination between XLA and DXS178 obtained by two point linkage analysis for clinical application. The physical extent of this region is not known but is likely to be in the order of 2–2.5Mb since we have now constructed a 2.5Mb YAC contig which spans the region between DXS366 and DXS101 (unpublished observations) and the distance between DXS366 and DXS442 is known to be less than 250kb[19]. The genetic and physical mapping of this region of Xq22 opens up the way for more directed cloning efforts for this gene and will allow the rapid assessment of candidate genes for XLA.

## MATERIALS AND METHODS

### Families

Blood samples were collected in EDTA for the preparation of DNA from individuals in 29 families affected by XLA. The affected boys in each family were identified as having XLA by the characteristic absence or very low circulating levels of all immunoglobulin isotypes and B lymphocytes. We included all the XLA families we had access to in the linkage analysis, assuming a mutation rate of $5\times10^{-6}$ for the XLA gene. In women where a recombination was observed, paternity was checked using the DXS255 hypervariable probe M27$\beta$[8].

### Probes

The probes used in this study were: p19-2 (DXS3), RX329 (DXS366), RX276 (DXS442), 212XT (DXS178), pKZO33H4 (DXS265), B1.4 (DXS101), pXG-12 (DXS94), pS21/S9 (DXS17) and 327M (DXS327) and are listed in Davies *et al.*[20], unless stated otherwise. We used the B1.4 probe to detect the DXS101 RFLP instead of the more commonly used cX52.5 probe. This 1.4kb *Bam*HI fragment was isolated from a cosmid which cross-hybridizes to several DXS101 species including the *Msp*I RFLP and detects less background species[6]. A 1kb *Xba*I/*Taq*I fragment of p212/9 was used to detect the DXS178 polymorphism. This fragment, 212XT, does not hybridize to the 9kb *Taq*I non-polymorphic fragment seen using the whole p212/9 probe. A 1.9kb *Msp*I fragment from pQST7H1 was used to detect the DXS327 RFLP. This probe, 327M, was preannealed for 3 hours with 2.5mg/ml human sonicated DNA at 65°C and washed to a final stringency of 0.5×SSC at 65°C. Cross-hybridization of this probe to repeat sequences was substantially reduced. All probes were labeled to high-specific activity with [$^{32}$P]-dCTP by the random-priming method[21].

### Linkage analysis

For DXS3 the results with two enzymes were combined into a haplotype and entered as a four allele system. Haplotype frequencies for DXS3 were calculated using data from unrelated individuals and were entered as follows: *Taq*I +, *Msp*I +,0.63; *Taq*I +, *Msp*I −,0.12; *Taq*I −, *Msp*I +,0.25; *Taq*I −, *Msp*I −,0.01.

Two point lod scores were calculated using the LIPED computer program; multi-point data were calculated using the ILINK and LINKMAP sections of the LINKAGE program (version 5.04)[11]. The gene frequency for XLA was entered as 1/100,000 and the mutation rate as $5\times10^{-6}$. The Haldane mapping function was used to calculate genetic distances.

## ACKNOWLEDGEMENTS

We would like to thank our many clinical colleagues for providing blood and/or DNA samples, in particular Dr T.Espanol, Dr G.Fontan, Dr P.Ioannou, Dr M.Schwartz, and Dr D.Webster, Dr D.Barker for providing probes, the Wellcome Trust and the Child Health Research Appeal Trust of the Institute of Child Health for their generous support.

## REFERENCES

1. Guioli,S., Arveiler,B., Bardoni,B., Notarangelo,L.D., Panina,P., Duse,M., Ugazio,A., Oberlé,I., de Saint Basile,G., Mandel,J.L. and Camerino,G. (1989) *Hum. Genet.*, **84**:19–21.
2. Kwan,S.-P., Terwilliger,J., Parmley,R., Raghu,G., Sandukyl,L.A., Ott,J., Ochs,H., Wedgwood,R. and Rosen,R. (1990) *Genomics*, **6**:238–242.
3. Kwan,S.-P., Terwilliger,J., Hagemann,T., Ott,J., Ochs,H., Wedgwood,R. and Rosen,F. (1991) *Cytogenet. Cell Genet.*, **58**:A27489.
4. Lovering,R., Sweatman,A.K., O'Reilly,M.-A.J., Genet,S., Middleton-Price,H., Malcolm,S., Levinsky,R.J. and Kinnon,C. *Hum. Genet.*, In Press.
5. O'Reilly,M.-A.J. (1991) PhD Thesis, University of London, London, UK.
6. O'Reilly,M.-A.J., Sweatman,A.K., Bradley,L.D., Alterman,L.A., Lovering,R., Malcolm,S., Levinsky,R. and Kinnon,C. in preparation.
7. Barker,D.F., Fain,P.R., Goldgar,D.E., Dietz-Band,J.N., Turco,A.E., Kashtan,C.E., Gregory,M.C., Tryggvason,K., Skolnick,M.H., Atkin,C.L. (1991) *Hum. Genet.*,**88**:189–194.
8. Boyd,Y. and Fraser,N.J. (1990) *Genomics*, **7**:182–187.
9. Arveiler,B., Oberlé,I. and Mandel,J.L. (1987) *Genomics*, **1**:60–66.
10. O'Reilly,M.-A.J., Alterman,L.A., Zijlstra,J., Malcolm,S., Levinsky,R. and Kinnon,C. *Genomics*, In Press.
11. Lathrop,G.M., Lalouel,J.M., Julier,C. and Ott,J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**:3443–3446.
12. Kwan,S.-P., Kunkel,L., Bruns,G., Wedgwood,R.J., Latt,S. and Rosen,.F.S (1986) *J. Clin. Invest.*, **77**:649–652.
13. Mensink,E.J.B.M., Thompson,A., Schot,J.D.L., Van De Greef,W.M.M., Sandukyl,L.A. and Schuurman,R.K.B. (1986) *Hum. Genet.*, **73**:327–332.
14. Malcolm,S., de Saint Basile,G., Arveiler,B., Lau,Y.L., Szabo,P., Fischer,A., Griscelli,C., Debre,M., Mandel,J.L., Callard,R.E., Robertson,M.E., Goodship,J.A., Pembrey,M.E. and Levinsky,R.J. (1987) *Hum. Genet.*, **77**:172–174.
15. Lau,Y.L., Levinsky,R.J., Malcolm,S., Goodship,J., Winter,R. and Pembrey, M. (1988) *Amer. J. Med. Genet.*, **31**:437–448.
16. Huang,T.H.-M., Cottingham,R.W., Ledbetter,D.H. and Zoghbi,H.Y. (1992) *Genomics*, **13**:375–380.
17. de Weers,M., Mensink,R.G., Kenter,M. and Schuurman,R.K. (1992) *Hum. Mol. Genet.*, **1**:653.
18. Allen,R.C. and Belmont,J.W. (1992) *Hum. Mol. Genet.*, **1**:216.
19. Sweatman,A.K., de Weers,M. Bradley,L.D., Lovering,R., O'Reilly,M.-A., Mensink,R.G.J., Hendricks,R.W., Schuurman,R.K.B., Levinsky,R.J. and Kinnon,C. in preparation.
20. Davies,K.E., Mandel,J.-L., Monaco,A.P., Nussbaum,R.L. and Willard, H.F. (1991) *Cytogenet. Cell Genet.*, **58**:853–966.
21. Feinberg,A.P. and Vogelstein,B. (1984) *Analyt. Biochem.*, **137**:266–267.

Hum Genet (1993) 91:605–608

human genetics



*Short communications*

# Isolation and mapping of discrete DXS101 loci in Xq22 near the X-linked agammaglobulinaemia gene locus

**Marie-Anne J. O'Reilly[1], Angela K. Sweatman[1], Linda D. Bradley[1], Lesley A. Alterman[1], Ruth Lovering[1], Sue Malcolm[2], Roland J. Levinsky[1], Christine Kinnon[1]**

[1] Molecular Immunology Unit, Institute of Child Health, London, UK
[2] Molecular Genetics Unit, Institute of Child Health, London, UK

Received: 24 August 1992

**Abstract.** The X-linked agammaglobulinaemia (XLA) gene locus has previously been mapped to Xq22 in genetic linkage studies. The DXS101 locus has shown no recombinations with XLA in the ten informative meioses investigated so far. The DXS101 sequence, recognised by the cX52.5 plasmid, is moderately repeated in Xq22. We have isolated cosmids which contain this sequence: two copies of which have been found to lie near DXS178 and XLA, and a third copy which lies near the PLP gene, distal to these loci. We have used the cosmids to generate probes which should be of use for RFLP analysis, and thus in both prenatal diagnosis and carrier testing for XLA, and in constructing a genetic map of this region. These probes will also be used to complement the genetic map in the construction of a complete physical map of Xq22.

## Introduction

X-linked agammaglobulinemia is a rare genetic disorder which has been previously mapped to the Xq22 region of the human X chromosome by genetic linkage studies (Malcolm et al. 1987; Guioli et al. 1989; Kwan et al. 1990). The likely order of loci in this region, given by these and other studies (O'Reilly 1991; Kwan et al. 1991; Lovering et al. 1993, and unpublished results), is centromere–DXS3–(DXS101, DXS178, DXS265, DXS327, DXS366, DXS442, XLA)–DXS94–DXS17–telomere. DXS178 has been reported to show no recombination with the XLA locus in over thirty informative meioses (Guioli et al. 1989; Kwan et al. 1990). DXS265, DXS327, DXS366 and DXS442 have shown no recombinations with the XLA locus in nine families (Kwan et al. 1991). DXS265 maps to within 5 kb of DXS178 and shows a significant degree of linkage disequilibrium with this marker (Lovering et al. 1993).

*Correspondence to:* C. Kinnon, Molecular Immunology Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK

DXS101 was originally reported as a polymorphic locus mapping to Xq21.3–22, but the DXS101-specific probe cX52.5 recognized at least five species, which all map to the same region of the X chromosome (Hofker et al. 1987). This probe has not been used extensively for genetic linkage analysis of XLA families. The *Msp*I-generated polymorphic alleles are difficult to resolve on standard agarose gels because of their similiarity in size (7.5 and 7.7 kb) and because several other irrelevant fragments are also detected (Hofker et al. 1987). We have recently investigated five XLA families and obtained a maximum LOD score of $Z = 3.0$ at $\theta = 0$ with 95% confidence limits of $0 < \theta < 0.2$ for DXS101 (O'Reilly 1991). One other genetic linkage study utilised DXS101 to investigate two Alport syndrome families and confirmed that the DXS101 polymorphism must lie in the region of the XLA locus as it mapped 1 cM proximal to DXS94 in multipoint analysis (Barker et al. 1991). In order to make this probe more useful for both genetic and physical mapping studies, we have isolated single copy sequences from the DXS101 plasmid and from two DXS101 positive cosmids. These probes should be useful for completing genetic and physical maps around the XLA locus.

## Materials and methods

### *Isolation and mapping of DXS101 positive cosmids*

The ICRFc100 X chromosome referece library (Nizetic et al. 1991) was screened with the DXS101 locus-specific probe cX52.5 and positive cosmids were selected. Four cosmids, ICRF D1219, F0476, GO220 and HO744 appeared to contain similar inserts, and a single representative cosmid, GO220, was analysed further. Cosmids EO928 and BO440 appeared to contain different inserts and were also chosen for further analysis.

### *PFGE and RFLP analysis*

All aspects of DNA preparation, conventional electrophoresis, PFGE and Southern blotting have been previously described

(O'Reilly et al. 1992, 1993). For RFLP analysis, genomic DNA was restricted with *Msp*I and size fractionated on 0.7% agarose gels electrophoresed under standard conditions for 2 days to separate the 7.5 and 7.7 kb alleles. Wash conditions for the various probes were as described below or in the figure legends.

### *DNA probes*

The DXS101 locus-specific plasmid cX52.5 contains a 3.6-kb *Eco*RI/*Pst*I insert which recognises multiple species in genomic Southern blot analysis (Hofker et al. 1987). A 1.1-kb *Sal*I/*Pst*I fragment, GPS, can be produced which detects single-copy species in this analysis. GPS hybridised very strongly to DNA from cosmid GO220 and extremely weakly to DNA from cosmids EO928 and BO440, indicating that the GO220 cosmid contains the DNA sequences corresponding to those in the cX52.5 plasmid. The B1.4 probe is a 1.4-kb *Bam*HI fragment which was subcloned from cosmid BO440. This fragment hybridised strongly with the DXS101 plasmid cX52.5 and is partially, but not completely, contained within the 3.6-kb insert, it is, however, distinct from the GPS fragment. It recognises a highly conserved fragment shared by all of the DXS101 species and was used for the cosmid mapping studies and RFLP analysis. A 700-bp *Bam*HI fragment from cosmid BO440 was also subcloned and sequenced to derive the following PCR primers, which yield a single 550-bp fragment following amplification of human genomic DNA, referred to as B550 in this study: primer 1: 5′ CTG AAG ACC ATG ATG GAC TAA GC; primer 2: 5′ CAG TGC ATA TGG GAT ACT CTG TAC.

Additional Xq22 specific probes used in this study were the DXS178 probe p212/9, the PLP locus-specific cDNA probe PRL1 and the DXS54 probe St3 (Davies et al. 1991). All of the probes were labelled to a high specific activity with [$^{32}$P]-dCTP using the random priming technique.

## Results

### *Mapping of the DXS101 sequences contained in the GO220 and EO928 cosmids*

Since the genetic linkage studies revealed that at least one copy of DXS101 must lie near DXS178, our initial aim was to map DXS101 relative to DXS178. The DXS178 probe p212/9 detects a 1.5-Mb partial *Mlu*I fragment (Fig. 1A, and O'Reilly et al. 1993) and a small 130-kb complete digest product (Fig. 1A). When the same filter was hybridised with the cX52.5 probe a smear between about 1.0 and 1.5 Mb was observed (Fig. 1A). Using the GPS fragment a single species of about 1.5 Mb was observed (Fig. 1A). That GPS and DXS178 lie on the same 1.5-Mb *Mlu*I fragment could be confirmed using the additional enzymes *Bss*HII (Fig. 1B) and *Nru*I (results not shown). GPS recognised a partial *Bss*HII fragment of 540 kb, in common with the DXS178 probe p212/9 (Fig. 1B), and a complete *Bss*HII digestion product of 80 kb, which is substantially smaller than the 460 kb fragment seen for DXS178 (Fig. 1B, and results not shown). This suggests that this copy of DXS101 lies more than 460 kb but less than 540 kb from the DXS178 locus.

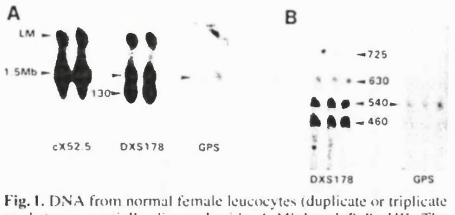



**Fig. 1.** DNA from normal female leucocytes (duplicate or triplicate tracks) was partially digested with: *A Mlu*I and *B Bss*HII. The DNA was electrophoresed and filters prepared. The filters were hybridised sequentially with the locus-specific probes (or loci) indicated. PFGE gel run conditions: *A* 40-V with a switch time of 2000 s for 180 h; *B* 120 V with a switch time of 120 s for 64 h. Molecular weight markers were *S. cerevisiae* (strain YP148) and *H. wingei* chromosomes and molecular weights of fragments detected are indicated in kb except where stated as Mb. LM = limiting mobility

The copy of DXS101 contained in the EO928 cosmid is likely to be different from that detected by the GPS fragment as revealed by fine mapping studies of the GO220 and EO928 cosmids using the rare cutting restriction enzymes *Bss*HII, *Nru*I, *Eag*I and *Nar*I (results not shown). Although both cosmids shared some sites in common not all were present in both cosmids, leading us to conclude that the DXS101 sequences contained in these two cosmids represent two distinct, yet similar, copies of these sequences. Attempts to isolate single copy sequences from the EO928 cosmid to aid further mapping of this sequence have been largely unsuccessful. However, it is likely that this copy of DXS101 also maps near DXS178 since a number of YAC clones isolated recently, which map to within 1 Mb of DXS178, have been found to contain this sequence (our unpublished observations).

### *Mapping of the DXS101 sequence contained in the BO440 cosmid*

The B550 probe from cosmid BO440 hybridised to a 1.06 Mb *Mlu*I fragment of normal leucocyte DNA, the same size as that recognised by probes specific for the Xq22 loci DXS24, DXS54 and DXS83 (Fig. 2B, C, and O'Reilly et al. 1993). When DNA from the 4X chromosome containing cell line LCL-127 was analysed in this way two fragments of 520 kb and 580 kb were detected with B550, the DXS54 probe St3 and the PLP locus-specific probe PRL1. The additional small fragment seen in the PRL1-probed tracks can be accounted for in that PRL1 is a cDNA probe. These data suggest that all three of these loci lie within 520 kb of each other. Additional mapping data, generated using *Bss*HII and *Sfi*I digests (Fig. 2), suggest that B550 and DXS54 are within 240 kb of each other and within 520 kb of PLP, but they cannot be orientated relative to each other.

### *The DXS101 MspI RFLP*

We have been able to overcome some of the practical problems faced when using the cX52.5 plasmid as a probe in genetic linkage studies by using the B1.4 fragment (Fig. 3). In an attempt to determine which sequence(s) is in the proximity of the *Msp*I polymorphic site, the cX52.5 positive cosmids were digested with *Msp*I and hybridised with cX52.5. None of the cosmids contained the 7.5 or 7.7 kb polymorphic *Msp*I fragments (results not shown).

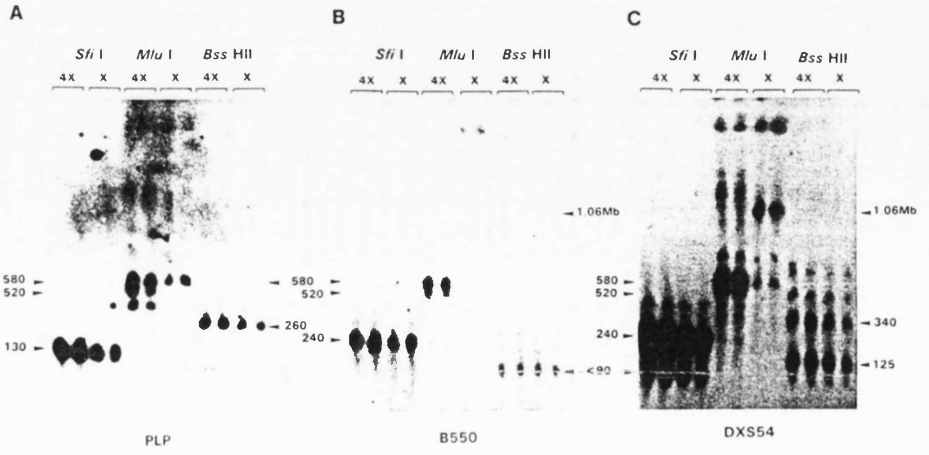



**Fig. 2.** Duplicate samples of DNA from LCL-127 cells (4X) or normal female leucocytes (X) was digested with the enzymes shown and size fractionated as 170 V with a switch time of 170 s for 18 h, then with a switch time of 70 s for 20 h. The filter was hybridised sequentially with the PLP, B550 and DXS54 locus-specific probes. All other details are as the legend to Fig. 1

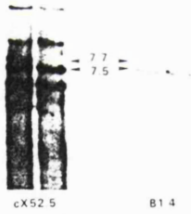



**Fig. 3.** Normale female DNA was investigated by Southern blot analysis for the DXS101 *Msp*I RFLP. The filter was hybridised sequentially with the radiolabelled cX52.5 and B1.4 probes and washed to a final stringency of 3×SSC and 1×SSC, respectively, at 65°C. The sizes of the polymorphic alleles are indicated in kb

## Discussion

In this report we describe the analysis of DXS101 positive cosmids and the isolation of single copy fragments in order to ascertain their proximity to each other and their position in Xq22. Two of these cosmids, EO928 and GO220, contain DXS101 sequences which appear to lie within about 1 Mb of DXS178 and should therefore be useful in further mapping this region of Xq22. The other copy of DXS101, contained in the BO440 cosmid and recognised by the single-copy probe B550, lies near the PLP gene locus, which is distal to DXS178 and XLA (O'Reilly et al. 1993). At present we cannot determine the position of other two proposed copies of this sequence in Xq22.

Although it could not be demonstrated that either of the cosmids mapping near DXS178 contain the *Msp*I polymorphic site, the polymorphism must presumably map near or within these copies of DXS101 because of the meioses investigated so far the polymorphic locus does not appear to recombine with the XLA locus or DXS178 (O'Reilly 1991). With the isolation of the B1.4 probe, which will make RFLP analysis for this locus more amenable, many more XLA families should now be investigated. Based on the allele frequencies (Hofker et al. 1987; O'Reilly 1991) 46% of females should be heterozygous for this marker and, therefore, this probe could be useful for genetic counselling in families which are not informative for the probes already in clinical use, namely DXS3, DXS17, DXS94, DXS178 and DXS265.

In this study we describe the characterisation of three of the estimated five copies of DXS101 which lie in Xq22. The isolation of additional cosmids from other libraries and more fine mapping studies of these regions may help to clarify this situation. In conclusion, the mapping of the various species of DXS101 and the isolation of single-copy sequences for these species should be useful for generating complete physical and genetic maps of Xq22 and for the eventual isolation of genes which map to this region of the X chromosome.

*Acknowledgements.* We would like to thank Dr. David Vetrie for discussing his unpublished results with us, the Wellcome Trust, the Lee Smith Research Fund, the Stephen Rutty Fund and the Child Health Research Appeal Trust of the Institute of Child Health for their generous support.

## References


Barker DF, Fain PR, Goldgar DE, Dietz-Band JN, Turco AE, Kashtan CE, Gregory MC, Tryggvason K, Skolnick MH, Atkin CL (1991) High density genetic and physical mapping of DNA markers near the X-linked Alport syndrome locus: definition and use of flanking polymorphic markers. Hum Genet 88:189–194

Davies KE, Mandel J-L, Monaco AP, Nussbaum RL, Willard HF (1991) Report of the Committee on the Genetic Constitution of the X Chromosome. Cytogenet Cell Genet 58:853–966

Guioli S, Arveiler B, Bardoni B, Notarangelo LD, Panina P, Duse M, Ugazio A, Oberlé I, de Saint Basile G, Mandel JL, Camerino G (1989) Close linkage of probe p212 (DXS178) to X-linked agammaglobulinemia. Hum Genet 84:19–21

Hofker MH, Bergen AAB, Skraastad MI, Carpenter NJ, Veenema H, Conner JM, Bakker E, van Ommen GJB, Pearson PL (1987) Efficient isolation of X chromosome-specific single-copy probes from a cosmid library of a human X/hamster hybrid-cell line: mapping of new probes close to the locus for X-linked mental retardation. Am J Hum Genet 40:312–328

Kwan S-P, Terwilliger J, Parmley R, Raghu G, Sandukyl LA, Ott J, Ochs H, Wedgwood R, Rosen R (1990) Identification of a closely linked DNA marker, DXS178, to further refine the X-linked agammaglobulinemia locus. Genomics 6:238–242

Kwan S-P, Terwilliger J, Hagemann T, Ott J, Ochs H, Wedgwood R, Rosen F (1991) Linkage mapping of closely linked DNA markers for X-linked agammaglobulinemia. Cytogenet Cell Genet 58:A27489

Lovering R, Sweatman AK, O'Reilly M-AJ, Genet S, Middleton-Price H, Malcolm S, Levinsky RJ, Kinnon C (1993) Physical mapping identifies DXS265 as a useful genetic marker for carrier detection and prenatal diagnosis of X-linked agammaglobulinemia. Hum Genet 91:178–180

Malcolm S, de Saint Basile G, Arveiler B, Lau YL, Szabo P, Fischer A, Griscelli C, Debre M, Mandel JL, Callard RE, Robertson ME, Goodship JA, Pembrey ME, Levinsky RJ (1987) Close linkage of random DNA fragments from Xq21.3–22 to X-linked agammaglobulinemia (XLA). Hum Genet 77:172–174

Nizetic D, Zehetner G, Monaco AP, Gellen L, Young BD, Lehrach H (1991) Construction, arraying, and high-density screening of large insert libraries of human chromosome X and 21: their potential use as reference libraries. Proc Natl Acad Sci USA 88:3233–3237

O'Reilly M-AJ (1991) Physical and genetic mapping of the X-linked agammaglobulinemia locus. PhD Thesis, University of London, UK

O'Reilly M-AJ, Alterman LA, Malcolm S, Levinsky RJ and Kinnon C. (1992) Identification of CpG-islands around the DXS178 locus in Xq22, near to the X-linked agammaglobulinemia locus. Hum Genet 90:275–278

O'Reilly M-AJ, Alterman LA, Zijlstra J, Malcolm S, Levinsky RJ, Kinnon C (1993) Pulsed field gel electrophoresis and radiation hybrid mapping analyses enable the ordering of eleven DNA loci in Xq22. Genomics 15:275–282

Human Molecular Genetics, 1994, Vol. 3, No. 1 79–83

# Mutation detection in the X-linked agammaglobulinemia gene, *BTK*, using single strand conformation polymorphism analysis

**Linda A.D.Bradley, Angela K.Sweatman, Ruth C.Lovering, Alison M.Jones, Gareth Morgan, Roland J.Levinsky and Christine Kinnon***

Molecular Immunology Unit, Division of Cell and Molecular Biology, Institute of Child Health, University of London, 30 Guilford Street, London WC1N 1EH, UK

Received September 9, 1993; Revised and Accepted November 19, 1993

**The gene defective in X-linked agammaglobulinemia (XLA) has recently been isolated and identified as *btk*, a non-receptor protein tyrosine kinase. We have utilized the technique of single strand conformation polymorphism (SSCP) analysis for the *btk* gene to identify mutations in XLA patients. The *btk* gene in affected boys from 10 families was analysed and mutations were identified in eight cases; seven of these were point mutations and one was a small insertion. The mutations were found throughout the gene coding region. Six of the patients have classical XLA and two have less severe forms of the disease. We have also identified a polymorphism at nucleotide position 2031. This technique will allow us to provide more accurate diagnoses of the disease and to determine the nature of the functional defects in the *btk* gene in these families.**

## INTRODUCTION

X-linked agammaglobulinemia (XLA) is a humoral immunodeficiency disease which results in developmental defects in the maturation pathway of B cells. Affected boys have normal levels of pre-B cells in their bone marrow but virtually no circulating mature B lymphocytes. This results in a lack of immunoglobulins of all classes and leads to recurrent bacterial infections. Treatment in most cases is by infusion of intravenous immunoglobulin.

The XLA locus had been mapped to Xq22 by genetic linkage studies[1–4] but until recently the defective gene was not known. The *btk* gene was isolated and suggested to be causative of XLA on the basis of its proximity to the closest genetic marker for this locus, the identification of mutations in 8 XLA patients and reduced levels of Btk protein and activity in patient cell lines[5,6]. The Btk protein appears to be a member of a new family of non-receptor protein tyrosine kinases related to, but distinct from, the Src family (reviewed in ref. [7,8]). Like Src, Btk has SH2 and SH3 domains and a catalytic domain with an ATP binding site[5,6].

Since XLA apparently results from the failure of pre-B cells to develop into functional B cells, the Btk protein is most likely to be involved in the B cell maturation process. Several protein tyrosine kinases have already been shown to be involved in the signalling pathways of lymphoid cells (reviewed in ref.[9]) and Btk may play a similar role within the signal transduction pathways of maturing B cells. However, the exact function of Btk has still to be established. In contrast to other disease-causing tyrosine kinase genes, XLA results from the lack of function of the Btk protein and therefore the XLA patients provide natural 'knockouts' of the *btk* gene. The characterisation of the *btk* gene in these patients should enhance our understanding of the functional domains of this protein and its role in B cell development.

We have already detected gene deletions in over 10% (6/50) of XLA patients investigated so far, providing a basis for genetic counselling in these families ([5]R.Lovering, in preparation). Three additional patients have been identified with altered restriction digest patterns in Southern blot analysis, two of whom were found to have point mutations[5]. In order to identify the nature of the mutations in the remaining majority of XLA patients we have used the technique of single strand conformation polymorphism (SSCP) analysis[10]. The *btk* cDNA was divided into seven sections each of which could be amplified by PCR. SSCP band shifts allowed the location of the altered base(s) to be identified to within approximately 200 base pairs. The simplicity of the technique gives it the benefit of being quickly transferable between laboratories and it can easily be used in clinical genetics services.

The SSCP technique was used to screen the *btk* gene in 10 unrelated XLA patients who were known not to have deletions or rearrangements in genomic DNA at this locus. Mutations were identified in 8 of these patients. Most patient phenotypes can be classified as 'classical' XLA having less than 1% B cells and undetectable immunoglobulin. In families with the classical phenotype all affected males have a similar severe phenotype. There are other less severe or 'leaky' XLA phenotypes where the patients have more than 1% B cells and detectable immunoglobulin; normal B cell levels are 10–20% of lymphocytes. In families with the leaky phenotype there is considerable heterogeneity in B cell numbers and immunoglobulin levels and thus, in the severity of the disease. Six of the patients studied have classical XLA and two have leaky forms of the disease.

## RESULTS AND DISCUSSION

Although XLA patients have no or few B cells, *btk* can usually be detected in cDNA prepared from peripheral blood leucocytes by PCR amplification. The patient cDNAs were amplified using primer pairs A to G (Table 1 and Fig. 3) sequentially and analysed using the SSCP technique. Sequence analysis of the fragments

* To whom correspondence should be addressed

which resulted in band shifts showed point mutations in seven of the patients and a 21 bp insertion in one patient.

### Premature stop codons in *btk* cause the classical XLA phenotype

Two of the patients were found to have single base pair changes resulting in premature stop codons within the *btk* gene. Patient AP was found to have an SSCP band shift in the smaller restriction digest fragment (112bp) from PCR reaction A at the 3′ end of the gene (Fig. 1a). Sequencing this region of the PCR product identified a mutation from G to T at nucleotide position 2038 resulting in a premature stop codon at amino acid residue 636. This mutation should lead to the loss of the terminal 24 amino acids from the protein, including several highly conserved residues. As there are three affected boys in this family who have no detectable B cells or immunoglobulin, this result suggests that the last 24 amino acids of this protein are critical for its correct expression and/or function in B cell development.

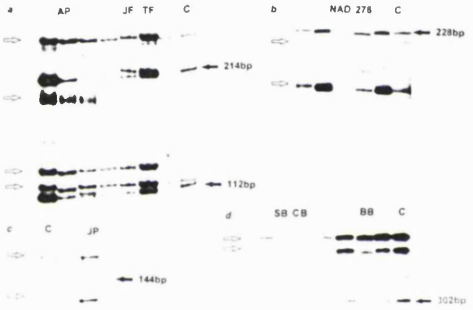



**Figure 1.** SSCP analysis of PCR products following digestion with the appropriate restriction enzyme for patient and control-(C) samples. Patients with mutations detected by SSCP are indicated by their initials, unlabelled tracks are patients that do not appear to have mutations in this analysis. Double stranded products are indicated by black arrows on the right and single stranded products indicated by white arrows on the left. (**a**) Reaction A showing single strand band shifts in affected brothers TF and JF in the 214 bp fragment and in patient AP in the 112 bp fragment. (**b**) Reaction E showing an altered banding pattern in patients NAD and 276. (**c**) Reaction D showing the altered single strand banding pattern in patient JP. (**d**) Reaction G showing the altered banding pattern in affected brothers SB and CB and in unrelated BB.

Patient NAD was found to have a band shift in the large fragment (228 bp) from PCR reaction E (Fig. 1b). On sequence analysis, he was found to have a base change of C to T at nucleotide position 895, resulting in a stop codon at amino acid residue 255 in the SH3 domain of Btk. This mutation should result in the production of a severely truncated protein lacking the remaining 404 amino acids, which includes all of the SH2 and kinase domains. This patient and his brother have no detectable B cells confirming that the absence of the functional domains of Btk results in a classical XLA phenotype.

### Mutations in the SH2 domain

Two patients were found to have point mutations in the SH2 domain of the Btk protein. Patient 276 is a member of a large Dutch XLA pedigree whose affected members exhibit heterogeneity in their clinical and immunological phenotypes[11]. Members of this family were found to have a deletion in the region of the XLA locus, at DXS442, co-segregating with the disorder, suggesting that the deletion may be causative of XLA in this family (A.K.Sweatman, in preparation). SSCP analysis showed band shifts in the 228 bp fragment resulting from the restriction digest of the product of PCR reaction E (Fig. 1b). Subsequent sequence analysis of this region showed a C to T change at nucleotide position 994 of the *btk* gene, which results in arginine being substituted for by a tryptophan at amino acid residue 288[11] (αA2, notation as in ref.[12]) and was confirmed in this laboratory. X-Ray crystallography and modelling of the Src SH2 domain suggests that this domain forms a two holed socket which can bind specific phosphorylated tyrosine residues in a two pronged conformation[13]. This mutation can be predicted to severely disrupt phosphotyrosine binding the SH2 domain.

Patient JP, diagnosed as classical XLA, had an SSCP band shift in the smaller fragment (144 bp) from the restriction digest of the product from PCR reaction D (Fig. 1c). Sequence analysis showed a mutation of A to G at nucleotide position 1051 which results in the substitution of a glycine for arginine B5 (notation for SH2 domain amino acid sequence as in ref.[13]) at amino acid residue 307. Arginine is a positively charged amino acid and at position B5 it is involved in the binding interactions at the base of the phosphotyrosine binding pocket. The change to a neutral glycine residue is highly likely to disrupt the binding potential of this region. This is consistent with the observation that

**Table 1.** PCR primers and products for SSCP analysis of *btk*

| Reaction | | PCR Primers | Position | Size | Enzyme | Cutting site | Products |
|---|---|---|---|---|---|---|---|
| A | A11F | (5′-CGGAAGTCCTGATGTATAGCA-3′) | 1808–2133 | 326 bp | *Rsa*I | 2022 | 214 bp |
| | E3R | (5′-CAAGAAGCTTATTGGCGAGC-3′) | | | | | 112 bp |
| B | B5[5] | (5′-TGTCAGATTTGCTGCTGAAC-3′) | 1568–1872 | 305 bp | *Hae*III | 1755 | 188 bp |
| | 11R | | | | | | 117 bp |
| C | A5[5] | (5′-GGTGAAGGAACTGCTTTGAC-3′) | 1294–1690 | 397 bp | *Eco*RI | 1454 | 161 bp |
| | 10R | | | | | | 236 bp |
| D | 849F | (5′-ATGTATGAGTGGTATTCCAAACAC-3′) | 964–1336 | 373 bp | *Ava*II | 1108 | 144 bp |
| | 7R | (5′GGTCCTTTGGATCAATTTCC-3′) | | | | | 229 bp |
| E | NTSH3 | (5′-ATGGGCTGCCAAATTTTGGAG-3′) | 619–1021 | 403 bp | *Hae*III | 794 | 175 bp |
| | 5R | (5′TTTAGCAGTTGCTCAGCCTG-3′) | | | | | 228 bp |
| F | A1F | (5′-TGTGTTGAAACAGTGGTTCC-3′) | 320–668 | 349 bp | *Hae*II | 460 | 140 bp |
| | 2R | (5′-CAGGTTTTAAGCTTCCATTC-3′) | | | | | 209 bp |
| G | Start F | (5′-AGCTACCTGCATTAAGTCAGG-3′) | 77–378 | 302 bp | | | 302 bp |
| | EX1R | (5′-CTCTTCTCGGAATCTGTCTTTC-3′) | | | | | |

substitution of the equivalent residue in the Abl SH2 domain (Arg 171) disrupts phosphotyrosine binding and reduces the transforming activity of this protein[14]. This patient has less than 1% B cells and undetectable immunoglobulin levels indicating that the replacement of this highly conserved arginine residue completely abolishes the functioning of Btk. Thus, two patients, JP and 276, show amino acid substitutions at conserved arginine residues in the SH2 domain of the protein resulting in different phenotypes.

### Kinase domain mutations can cause the leaky XLA phenotype

Family F has three affected boys with B cell counts varying from less than 1% to 3%. Immunoglobulin levels vary considerably between the affected boys; TF has normal IgG, IgA, IgM, IgE and IgG subclass levels except for IgG2 which was 0.55 g/l (normal range 0.7–5.15g/l), JF has low normal IgG, 0. 16g/l IgA (normal range 0.3–1.3g/l), 0.22g/l IgM (normal range 0.5–2.2g/l), normal IgE and IgG subclasses except for IgG1 which was 2.4g/l (normal range 2.85–6.8g/l) and KF has normal IgG (IgG subclass levels not available) and IgA and 0.32g/l IgM (normal range 0.5–2.2g/l). Initially this family was diagnosed as having an undescribed B cell immunodeficiency but subsequently, on the basis of a non-random pattern of X-inactivation in the B cells of the obligate carrier female, they were diagnosed as XLA[15]. Two affected brothers, TF and JF, were studied using SSCP and found to have band shifts in the larger restriction fragment (214 bp) generated from PCR reaction A (Fig. 1a). The C to A mutation identified at nucleotide position 1952 causes a non-polar to polar amino acid substitution, alanine to aspartic acid, at amino acid residue 607 near the 3′ end of the gene. In Src and other related proteins this residue is not highly conserved but always appears to be non-polar. This amino acid substitution therefore disrupts but does not completely abolish the functioning of this protein probably by a reduction in kinase activity, as the change is in the kinase domain. The mutation in these patients confirms the XLA diagnosis and suggests that other patients with similar mild phenotypes may have mutations in the *btk* gene. Partial disruption of the functioning of Btk in cases such as this will provide a valuable insight into the structure/function relationship of this protein.

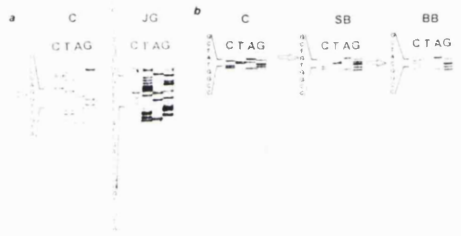



**Figure 2.** Autoradiographs of sequencing gels showing base changes compared to normal control (C) (**a**) as a result of a 21bp insertion in patient JG at nucleotide position 442 (position of insertion indicated in the normal control by a white arrow) and (**b**) in the start codon in patients SB and BB at nucleotide positions 133 and 134, respectively.

### An insertion in the 5′ end of *btk*

Patient JG has two affected brothers and all have undetectable B cells and immunoglobulin levels. An altered *Msp*I restriction fragment was detected on Southern blot analysis of this family as reported earlier (patient C, ref.[5]). No other alterations were observed on digestion with additional enzymes, suggesting no deletion or gene rearrangement. Subsequent *Msp*I digestion of the cDNA from this patient showed an altered restriction digest pattern and SSCP analysis produced a band shift in the smaller fragment (140 bp) from reaction F (results not shown). Sequence analysis showed that there is an insertion of 21 bp (T-CTGTGTTTTCATCGACCCGG) at nucleotide position 442 (amino acid 103) creating a new *Msp*I enzyme site (Fig. 2a). The origin of the extra sequence is not known but the insertion appears to occur at the junction of exons 3 and 4[16] and may therefore be due to a point mutation at a splice site in the intron. This extra DNA sequence results in the insertion of 7 amino acids (Ser-Val-Phe-Ser-Ser-Thr-Arg) between amino acids 103 and 104 in the protein. As the insertion is 21 bp, the rest of the protein

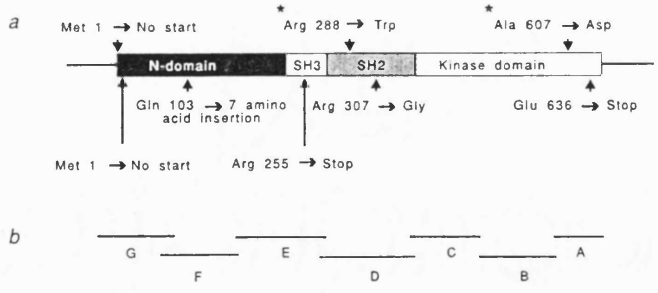



**Figure 3.** Schematic representation of the Btk protein showing (**a**) the approximate locations of the mutations and the amino acid changes and (**b**) the approximate extent of the PCR reaction products used for the SSCP analysis. * Indicates patients with 'leaky' XLA phenotype.

is still translated in frame and normal SH2, SH3 and kinase domains will be present.

The N terminal of this protein does not yet have a defined function, but a point mutation at nucleotide 216 changing amino acid residue 28 in the mouse *btk* gene results in the *xid* phenotype[16, 17] demonstrating the importance of the N terminal region in B cell development. The presence of the 7 extra amino acids seems to completely inactivate the protein as the affected males in this family have the classical XLA phenotype. This inactivation could be a result of a specific disruption of the N terminal region or it could be the result of a more general change in protein folding. Studies of the mutant protein should help to elucidate the exact nature of this defect.

### Initiation codon mutations

Patient BB was found to have an SSCP band shift in the 302 bp fragment from reaction G (Fig. 1d) and sequence analysis showed that he has a T to C point mutation in the start codon at nucleotide 134 (Fig. 2b). Another patient, SB, was found to have a band shift in the same fragment (Fig. 1d) and also has a point mutation in the start codon at nucleotide 133, changing A to G (Fig. 2b). Both of these patients have less than 1% B cells suggesting that the disruption of the initiation codon results in the lack of any functional Btk protein. It is possible that these mutations may result in the production of a smaller molecular weight protein beginning at one of the subsequent start codons since the initiation codons at nucleotide positions 317 and 619 have good Kozak consensus sequences. It is not yet known, however, if these initiation codons are functional.

Patient BB is a sporadic case and was initially diagnosed as having XLA with growth hormone deficiency[18, 19] but further studies have shown him to be an XLA patient of short stature. XLA with growth hormone deficiency is a very rare syndrome and it is difficult to be sure that the conditions are being inherited together unless genetic linkage analysis can be performed on the family. This study shows that caution should be exercised before diagnosis.

### Location of an SSCP polymorphism

We have found an SSCP polymorphism which produces different banding patterns of the 112 bp fragment from PCR reaction A. Sequence analysis of this fragment showed that the SSCP polymorphism was due to a C to T substitution at nucleotide 2031, which does not affect the amino acid sequence of the protein. The rarer allele has the cDNA sequence published for *btk*[5], T at nucleotide position 2031, and is present in four out of thirteen XLA families and four out of twenty-three leukemia patients[20]. The more common allele has a C at this position and is observed in 28 out of the 36 patients. The predicted heterozygosity of this polymorphism is 35%.

## GENERAL DISCUSSION

Two point mutations in *btk* have been described previously[5], both of which resulted in amino acid substitutions at highly conserved amino acid residues in the kinase domain, one within the substrate specific domain and one within the ATP binding site. We present here mutations in the *btk* gene found in eight unrelated XLA patients detected using SSCP analysis. These mutations are all different and spread throughout the gene (Fig. 3). So far, amino acid substitutions or insertions have been identified in all domains of the Btk protein, except the SH3 domain. No mutations were detected by SSCP in two patients both of whom are sporadic cases with a classical XLA phenotype. It is possible that the mutations in these boys were not detectable using this technique but it is also possible that their mutations are outside the coding sequence, in control or promoter elements. Alternatively, as they are both sporadic cases and X-inactivation analysis has not been performed to confirm X-linked inheritance, it is possible that their diagnoses as XLA are incorrect.

The identification of these mutations has important implications for genetic counselling for these families. Results of preliminary studies show that it is possible to determine carrier status using SSCP analysis when there is a known mutation (results not shown). In cases where a mutation has not been identified, the polymorphism described here may be useful as it has a predicted heterozygosity of 35%. This polymorphism can be detected by SSCP but does not involve a restriction enzyme site. Prenatal diagnosis based on such point mutations within the *btk* gene will become available when the genomic organisation of the gene is determined.

It appears that mutations in the *btk* gene are responsible for a range of B cell immunodeficiency disorders, from complete agammaglobulinemia to conditions where the patient has reduced amounts of B cells and immunoglobulin. It is possible that even milder phenotypes will be the result of less functionally important mutations in this gene. All the mutations found should provide valuable clues to the normal functioning of this protein in the pathway of B cell development.

## MATERIALS AND METHODS

### Patients

The patients are all males diagnosed as having XLA because of the complete absence or very low levels of circulating mature B lymphocytes. In some cases diagnosis was confirmed by X-inactivation studies in carrier females[15]. Other blood cell lineages appear normal. All patients were previously investigated for potential *btk* gene deletions or other rearrangements by Southern blot analysis of genomic DNA and were found to be comparable to normal.

### cDNA preparation

Lymphocytes were prepared from fresh blood by Ficoll (Pharmacia) separation and washed twice in PBSA. RNA was made from $2 \times 10^6$ cells, as described[21], and resuspended in 18 μl of water, before heating to 70°C for 5 min and then chilling on ice. 1 μl RNAGuard (Pharmacia), 2.5 μl 0.1 M dithiothreitol, 10 μl pd(T)$_{12-18}$ 0.1 mg/ml (Pharmacia), 10 μl 5 × buffer (BRL), 5 μl 10 mM dNTP, 2.5 μl bovine serum albumin 2 mg/ml and 2 μl Moloney murine leukemia virus reverse transcriptase 20 U/μl (BRL) were added in order and the sample incubated at 42°C for 90 min. The enzyme was then inactivated at 65°C for 5 min and the cDNA stored at −20°C.

### PCR amplification and SSCP analysis

The *btk* gene was amplified from cDNA using seven overlapping PCR reactions A to G, as shown in Table 1. The reaction was performed using 200 μM dGTP, dATP and dTTP and 20μM dCTP with 1 μCi [$^{32}$P] dCTP, 1.5 mM $MgCl_2$, 50 pmol of each primer (forward and reverse), buffer and enzyme according to manufacturer's instructions (Bioline) and 1 μl cDNA in a 50 μl final volume. The reactions were denatured at 94°C for 3 min and then given 30 cycles at 57°C for 30 seconds, 72°C for 30 seconds and 94°C for 30 seconds. The products were digested as indicated to give the fragment sizes shown (Table 1) by the addition of 10 units of enzyme into the PCR mix and incubation at 37°C overnight. The PCR products were analysed on 1% agarose gels and diluted in 0.1% SDS, 10 mM EDTA to between 1/2 and 1/40 to ensure even loading on the polyacrylamide gels.

2 μl of sample and 2 μl loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol) were denatured at 95°C for 3 min, loaded onto a 6% non-denaturing polyacrylamide gel and electrophoresed in 1 × TBE buffer for 5 h. An undenatured sample was always analysed on the same gel to allow identification of the double stranded molecules in the patient samples. Gels were run at 30 Watts at room temperature using 5% glycerol and at 60 Watts at 4°C using 10% glycerol. Both run conditions were used for each sample.

All band shifts were seen at both temperatures but the band shifts were more significant at 4°C. All SSCP gels shown in the figures were run at 4°C. Gels were dried and exposed to autoradiograph film (Kodak) for 3 days at room temperature.

**Sequence analysis**

The PCR products which resulted in identifiable band shifts on SSCP analysis were sequenced using template amplified with a biotinylated PCR primer StartF and primer A8R (5'-TCATCTTCAGACATGGAGCC-3') and/or ERB2 (biotinylated) (5'-ATTGAGTGGGAGCACAAAGG-3') and A5i. The PCR product was generated as for the SSCP analysis but using a 1 min 72°C extension time. The PCR product was separated on a 1% low melt point agarose gel, excised and purified using Geneclean Kit (Bio101). The sample was then attached to Dynabeads M-280 Streptavidin (Dynal) according to the manufacturer's instructions and the strands denatured. Sequencing was performed using a Sequenase Kit (USB). cDNA from patient NAD was sequenced by using 200 ng of startF/A8R PCR product in a Taquence Cycle Sequencing reaction (USB).

## ACKNOWLEDGEMENTS

We would like to thank all the health professionals who supplied us with patient samples, in particular Mike Tarlow, Adam Finn, Kevin Forsyth and Sheila Cochrane. Michel de Weers for providing the 276 patient cell line, Paul Baird for help in setting up the SSCP analysis and David Vetrie for communicating btk primer sequences prior to publication, the Wellcome Trust and the Child Health Research Appeal Trust of the Institute of Child Health for their generous support.

## REFERENCES

1. Kwan,S.-P., Kunkel,L., Bruns,G., Wedgwood,R.J., Latt,S. and Rosen,F.S. (1986) *J. Clin. Invest.* **77**, 649–652.
2. Malcolm,S., de Saint Basile,G. de, Arveiler,B., Lau,Y.L., Szabo,P., Fischer,A., Griscelli,C., Debre,M., Mandel,J.L., Callard,R.E., Robertson,M.E., Goodship,J.A., Pembrey,M.E. and Levinsky,R.J. (1986) *Hum. Genet.* **77**, 172–174.
3. Lovering,R., Middleton-Price,H.R., O'Reilly,M.-A.J., Genet,S.A., Parkar,M., Sweatman,A.K., Bradley,L.D., Alterman,L.A., Malcolm,S., Morgan,G., Levinsky,R.J. and Kinnon,C. (1993) *Hum. Mol. Genet.* **2**, 139–141.
4. Parolini,O., Hejtmancik,J.F., Allen,R.C., Belmont,J.W., Lassiter,G.L., Henry,M.J., Barker,D.F. and Conley,M.E. (1993) *Genomics* **15**, 342–349.
5. Vetrie,D., Vorechovsky,I., Sideras,P., Holland,J., Davies,A., Flinter,F., Hammarstrom,L., Kinnon,C., Levinsky,R., Bobrow,M., Smith,C.I.E. and Bentley,D. (1993) *Nature* **361**, 226–233.
6. Tsukada,S., Saffran,D.C., Rawlings,D.J., Parolini,O., Allen,R.C., Klisak,I., Sparkes,R.S., Kubagawa,H., Mohandas,T., Quan,S., Belmont,J.W., Cooper,M.D., Conley,M.E. and Witte,O. (1993) *Cell* **72**, 279–290.
7. Desidero,S. (1993) *Nature* **361**, 202–203.
8. Kinnon,C., Hinshelwood,S., Levinsky,R.J. and Lovering,R.C. (1993) *Immunol. Today* **14**, 554–558.
9. Perlmutter,R.M., Levin,S.D., Appleby,M.W., Anderson,S.J. and Alberola-Ila,J. (1993) *Annu. Rev. Immunol.* **11**, 451–499.
10. Orita,M., Iwahana,H., Kanazawa,H., Hayashi,K. and Sekiya,T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2766–2770.
11. de Weers,M., Mensink,R.G.J., Kraakman,M.E.M., Schuurman,R.K.B. and Hendriks,R.W. (1994) *Hum. Mol. Genet.* this issue.
12. Eck,M.J., Shoelson,S.E. and Harrison,S.C. (1993) *Nature* **362**, 87–91.
13. Waksman,G., Shoelson,S.E., Pant,N., Cowburn,D. and Kuriyan,J. (1993) *Cell* **72**, 779–790.
14. Mayer,B.J., Jackson,P.K., Van Etten,R.A. and Baltimore,D. (1992) *Mol. Cell. Biol.* **12**, 609–618.
15. Alterman,L.A., de Alwis,M., Genet,S., Lovering,R., Middleton-Price,H., Morgan,G., Jones,A., Malcolm,S., Levinsky,R.L. and Kinnon,C. (1993) *J. Immunol. Meth.* **166**, 111–116.
16. Rawlings,D.J., Saffran,D.C., Tsukada,S., Largaespada,D.A., Grimaldi,J.C., Cohen,L., Mohr,R.N., Bazan,J.F., Howard,M., Copeland,N.G., Jenkins,N.A. and Witte,O.N. (1993) *Science* **261**, 358–361.
17. Thomas,J.D., Sideras,P., Smith,C.I.E., Vorechovsky,I., Chapman,V. and Paul,W.E. (1993) *Science* **261**, 355–358.
18. Conley,M.E., Berks,A.W., Herrod,H.G. and Puck,J.M. (1991) *J. Pediat.* **119**, 392–397.
19. Fleisher,T.A., White,R.M., Broder,S., Nissley,S.P., Blaese,R.M., Mulvihill,J.J., Olive,G. and Waldmann,T.A. (1980) *New Engl. J. Med.* **302**, 1429–1433.
20. Katz,F.E., Lovering,R.C., Bradley,L.A.D., Rigley,K.P., Brown,D., Cotter, F., Chessels,J.M., Levinsky,R.J. and Kinnon,C. (1994) *Leukemia*, In Press.
21. Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.* **162**, 156–159.

# Expression of the X-Linked Agammaglobulinemia Gene, *btk* in B-Cell Acute Lymphoblastic Leukemia

Fay E. Katz[1], Ruth C. Lovering[1], Linda A.D. Bradley[1], Kevin P. Rigley[2], Derek Brown[2], Finbarr Cotter[1], Judith M. Chessells[1], Roland J. Levinsky[1], and Christine Kinnon[1]

[1]Molecular Immunology Unit, [2]Haematology and Oncology Unit, Institute of Child Health, London, and [2]Celltech, Slough, Bucks, UK

**The gene which causes X-linked agammaglobulinemia, *btk*, has recently been identified as a cytoplasmic tyrosine kinase expressed almost exclusively in B cells, and at all stages of B-cell differentiation. To assess the possibility of involvement of this gene in childhood B-cell malignancies, cells from 23 pediatric patients with B-cell acute lymphoblastic leukemia were examined for expression and alteration of the Btk protein and also for mutations in the *btk* gene. Btk proteins, similar in both molecular weight and quantity to those seen in unaffected individuals, were detected in whole cell lysates from the blasts of 12/12 patients indicating that no abnormal protein was present. cDNAs from the leukemic blasts of all 23 patients were screened with specific primers covering the coding region of the *btk* cDNA for mutations using single strand conformation polymorphism (SSCP) analysis. No mutations were found but a nucleotide polymorphism was identified in 4/23 patients at the 3' end of *btk*. Although the sample size in this study was relatively small, these data suggest that *btk* does not appear to play a critical role in childhood B-cell leukemias.**

## INTRODUCTION

Acute leukemia, though a rare disease, is the commonest form of childhood cancer. This is an extremely heterogeneous group of disorders, both at clinical and biological levels, and numerous agents or mechanisms have been implicated in the disease process. Both genetic and environmental factors have been associated with the development of childhood acute lymphoblastic leukemia (ALL) (1). Infection, either by bacteria or viruses *in utero* or early infancy has been suggested as a strong risk factor in the etiology of leukemia (2). In particular, viruses may activate proto-oncogenes which can then transform cells from the normal to the malignant phenotype.

X-linked agammaglobulinemia (XLA) is an inherited immunodeficiency disorder which manifests itself in affected boys by severely depressed or absent levels of all classes of serum immunoglobulins (3). Mature B cells are either absent altogether, or present in very reduced numbers and patients readily succumb to life-threatening infections (4). Using a positional cloning approach, Vetrie *et al.* (5) have recently isolated the *btk* gene (also known as *atk* or *bpk*) which maps to Xq22, the region known to contain the XLA gene (6,7). The expression of *btk* was shown to be restricted to the B-cell lineage and some myeloid cell lines. Deletions and point mutations were demonstrated in 8/33 XLA patients, suggesting that mutations in *btk* are causative of XLA (5). At the same time, Tsukada *et al.* (8) reported the isolation of a novel tyrosine kinase from a mouse pre-B cell cDNA library which is also expressed in B and myeloid cells. Sequence comparison showed this to be the murine homolog of *btk*.

Btk shares some structural features with members of the Src family of protein tyrosine kinases (5,8), although it lacks the C terminal regulatory tyrosine and N terminal myristylation sequence, and more closely resembles the two recently described cytoplasmic tyrosine kinases Tec and Itk (9,10). However, because of its similarity to the Src family and its expression at all stages of B-cell maturation, it was suggested that Btk may have an oncogenic potential in B cells (5). In two types of leukemia, namely Philadelphia-positive common ALL and chronic myeloid leukemia (CML), deregulation of tyrosine kinase activity has been shown to be actively involved in the disease process (11). The t(9;22), which is present in over 90% of cases of CML and 20% of cases of adult ALL, gives rise to a bcr-*abl* fusion gene and the resultant chimeric mRNA produces a hybrid protein with its N terminal derived from bcr (on chromosome 9) and its C terminal from *abl* (on chromosome 22). Deregulation of the tyrosine kinase activity of the normal *abl* product results. Other gene fusions, and by analogy fusion transcripts and proteins, have been shown to be implicated in acute leukemias carrying the t(1;19), t(6;9), t(15;17), t(8;14), and t(1;14) (12). In all of these cases, however, transcription factors rather than tyrosine kinases are involved in the genesis of the malignancy.

In order to investigate the possibility that *btk* plays a role in B-cell malignancy, we have examined a total of 23 cases of childhood B lineage ALL for expression of *btk*. These leukemias are thought to be representative of the early stages of normal B-cell development. Translocations, deletions or numerical changes involving the X chromosome are very rare in B lineage lymphoblastic leukemias (13), and none of the patients we studied had any obvious X chromosome-associated karyotypic abnormalities. Expression was investigated in 12 cases using a novel antiserum raised against the Btk protein. Point mutations in the *btk* gene were screened for in these, and a further 11 cases, using single strand conformation polymorphism (SSCP) analysis. Any mutations identified in this way could be implicated in the malignant transformation process.

## MATERIALS AND METHODS

### *Patient Samples*

Cryopreserved presentation leukemic blasts from 23 patients with B-cell ALL were studied. These comprised 11 patients with null (B precursor) ALL, five patients with common ALL, four patients with pre-B ALL, and three patients with mature B ALL. None of the patients studied had any detectable cytogenetic abnormalities involving the X chromosome. Clinical details of the patients are shown in Table 1.

Received August 13, 1993. Accepted November 9, 1993.
**Correspondence to**: Dr Christine Kinnon, Molecular Immunology Unit, Institute of Child Health, 30 Guilford St, London, WC1N 1EH, UK.

**Table 1** Clinical Details and Immunophenotypes of Leukemic Blasts from 23 Patients Investigated in This Study

| Patient | Age | Sex | WBC ($\times 10^9$/l) | FAB | Immunophenotype |
|---|---|---|---|---|---|
| 1 | 1.5 mo | M | 220 | L1 | (Null ALL) |
| 2 | 18 mo | F | 141 | L1 | CD10−, CD19+, HLADR+, cyt$\mu$−, TdT+/− |
| 3 | 12 mo | M | 370 | L1 | |
| 4 | 4 mo | F | 800 | L1 | |
| 5 | 1.5 mo | F | 760 | L1 | |
| 6 | 3 mo | F | 280 | L1 | |
| 7 | 3 mo | M | 180 | L1 | |
| 8 | 6 mo | F | 426 | L1 | |
| 9 | 8 mo | M | 190 | L1 | |
| 10 | 0.5 mo | M | 1018 | L1 | |
| 11 | 3 mo | M | 71 | L1 | |
| 12 | 3 yr | M | 25 | L1 | (Common ALL) |
| 13 | 6 yr | F | 77 | L1 | CD10+, CD19+, HLADR+, cyt$\mu$−, TdT+ |
| 14 | 6 yr | M | 188 | L1 | |
| 15 | 5 yr | M | 751 | L1 | |
| 16 | 4 yr | F | 136 | L1 | |
| 17 | 3 yr | F | 12 | L1 | (Pre B ALL) |
| 18 | 12 yr | M | 17 | L1 | CD10+, CD19+, HLADR+, cyt$\mu$+, TdT+ |
| 19 | 13 yr | M | 90 | L1 | |
| 20 | 5 yr | M | 3 | L1 | |
| 21 | 5 yr | M | 7 | L3 | (B ALL) |
| 22 | 5 yr | M | 9 | L3 | CD10−, CD19+, HLADR+, cyt$\mu$−, TdT−, SmIg+ |
| 23 | 4 yr | M | 22 | L3 | |

### *Antiserum Production*

Two oligonucleotide primers were designed to create a *Bam*HI-*Eco*RI polymerase chain reaction (PCR) amplification product containing the cDNA sequence for amino acids 163 to 218 of the Btk protein. This fragment was cloned and expressed in the bacterial expression plasmid PGEX-2T (14). The soluble fusion protein was purified (14), shown to be a single species of approximately the correct molecular weight by SDS-PAGE, and used to immunize NZ white rabbits as described by Tsukada *et al.* (8), except that the animals were boosted monthly. Antisera were collected at monthly intervals, starting 1 week after the first booster. The antisera were shown to be specific for the Btk protein by ELISA, Western blot analysis and immunoprecipitation (Lovering *et al.*, unpublished observations). By Western blot analysis, the 77 kDa protein identified specifically with the antiserum in B cells and peripheral blood lymphocytes was not present in T cells nor in the peripheral blood lymphocytes of XLA patients. Two additional species of 100 kDa and 70 kDa cross-reacted with

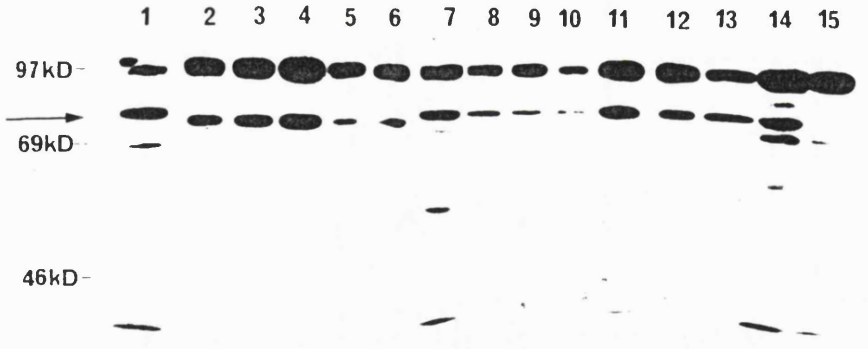



**Figure 1** Western blot analysis of leukemic whole cell lysates detected with anti-Btk antiserum. The tracks were loaded as follows: tracks 1, 2 and 7 null ALL, tracks 3, 4, 8, 9 and 10 common ALL, tracks 5, 6 and 11 pre-B ALL, track 12 B-ALL, track 13 normal peripheral blood mononuclear cells, track 14 B-cell line IY, track 15 T-cell line Jurkat. The 77 kDa Btk protein is indicated by an arrow. Equal cell numbers were loaded in each track and similar levels of protein were detected in the leukemic pre-B cell line Nalm-6 (results not shown).


**Figure 2** SSCP analysis of the 112 bp PCR product from the 3' end of the *btk* gene (nucleotide residues 2021 to 2133) from cDNAs of two leukemia patients (tracks 1 and 2), showing the nucleotide polymorphism at position 2031. The arrows to the left indicate single-stranded DNA of the rarer allele in track 1, while the arrows on the right indicate the single-stranded DNA of the more common allele in track 2. The unidentified species in track 1 is residual undenatured double-stranded DNA.

the antisera on Western analysis, however they were not immunoprecipitated by the antisera (results not shown). The pattern of expression of the 77 kDa and additional proteins was comparable directly with that identified by the anti-Bpk antisera generated by Tsukada *et al.* (results not shown).

### *Western Blot Analysis*

Whole cell lysates were prepared from cryopreserved leukemic blasts, where available, by hypotonic lysis in buffer (10 mM Tris pH 7.4, 2 mM $MgCl_2$, 1 mM $CaCl_2$) containing 1 mM PMSF (as a protease inhibitor) at a cell concentration of $5 \times 10^7$/ml. After lysis, an equal volume of reducing loading buffer was added to each sample, samples were then stored at −20°C. A volume of lysate corresponding to $1 \times 10^6$ cells was boiled for 5 min and then separated on 7.5% denaturing polyacrylamide gels in Tris/glycine/SDS buffer overnight at 40 V and room temperature. The electrophoresed proteins were transferred onto Hybond C nylon membranes (Amersham International, Amersham, UK), and stored at −20°C. Immediately prior to immunoblotting, the membranes were incubated in 5% non-fat milk in phosphate-buffered saline 'A' (PBSA) with gentle rocking for at least 2 h at room temperature to block non-specific binding. They were then rinsed briefly in 0.1% Tween 20/PBSA (TPBSA), and incubated for 1 h with rocking in 3% BSA/PBSA. After a second wash, they were incubated with a 1 : 100 dilution of purified anti-Btk antiserum in (200 mM Tris/0.05% Tween 20/1 fetal calf serum (TTF)) for at least 3 h with gentle rocking. The membranes were then washed twice for 5 min in TPBSA and then incubated with a 1 : 1000 dilution of horse-radish peroxidase conjugated goat anti-rabbit Ig antiserum in TTF for a further 1.5 h at room temperature. Following a further five washes in TPBSA, the antigen/antibody reactions were visualized using the enhanced chemiluminescence (ECL) detection system (Amersham International). The membranes were autoradiographed for time periods ranging from 30 s to overnight.

### *mRNA and cDNA Preparation*

mRNA was prepared from the cryopreserved leukemic cells using the FAST-TRACK kit (Invitrogen; British Biotechnology Products, Abingdon, UK) and used to prepare cDNA by a modified RT-PCR method incorporating random hexamer oligonucleotide primers. Normal B-cell mRNA and cDNA for use as controls were prepared from peripheral blood mononuclear cells and EBV transformed B-cell lines.

### *PCR Amplification and SSCP Analysis*

Overlapping primer sets were designed to span the entire coding region of the *btk* cDNA. Details of these primers are reported elsewhere (15). All cDNA samples were checked for reactivity with each primer set before SSCP analysis. PCR amplification was carried out on 1–3 μg cDNA using a Hybaid thermal reactor and the following conditions: an initial denaturation step of 94°C for 3 min, 30 cycles of 94°C (30 s), 57°C (30 s), 72°C (1 min), final extension step of 72°C for 10 min [$^{32}$P]-dCTP was added to the reaction mixture immediately prior to PCR amplification. PCR products greater than 250 bp were digested at 37°C overnight using an appropriate restriction enzyme (15). Digested and undigested PCR products were checked for adequate amplification and digestion before SSCP analysis.

### *SSCP Analysis*

Aliquots of the PCR products were diluted into a mixture of 0.1% SDS/10mM EDTA and 2 μl of each dilution added to an equal volume of stop solution (95% deionized formamide/20 mM EDTA/0.05% bromophenol blue/0.05% xylene cyanol). The samples were then denatured at 95°C for 5 min, cooled on ice and immediately loaded onto a 6% non-denaturing polyacrylamide gel (Acugel; National Diagnostics, Atlanta, GA, USA), containing 5% glycerol. Gels were routinely run in 1 × TBE buffer for 5–6 h, at 30 W, at room temperature and were then dried and autoradiographed for up to 4 days.

## RESULTS AND DISCUSSION

The 77 kDa Btk protein was detected in whole cell lysates from 12 patients studied; three cases of null ALL (patients 1, 2 and 9), five cases of common ALL (patients 12, 13, 14, 15 and 16), three pre-B ALLs (patients 17, 18 and 19), and one B ALL (patient 21) (Table 1). The majority of the retrospective group of patients had high white cell counts and extensive disease, representing the worst prognostic group in childhood ALL (16). No significant quantitative or qualitative differences in the Btk protein were seen in any patient group; the level of expression and the molecular weight of the protein, 77 kDa, appeared similar in all cases to that of peripheral blood mononuclear cells from unaffected individuals and B-cell lines (Figure 1).

Using RNA which had been isolated at presentation from all 23 patients, SSCP analysis of the prepared cDNAs (using primer pairs spanning the entire coding region of the *btk* gene) showed no band shifts associated with a mutation in the gene. Although the sample size studied here is relatively small, it appears unlikely that mutations in *btk* represent a major causative factor in B-cell leukemia. Band shifts were seen in four patients (patients 4, 6, 14 and 18), using the most 3' set

of primer pairs, and these were subsequently shown to be due to a polymorphism (Figure 2). This polymorphic band shift was also seen in 4/13 patients with XLA, giving an overall heterozygosity of 35%, and suggesting that it may be useful for the determination of XLA carrier status in some females (15).

Although no karyotypic abnormalities are associated with XLA, deletions, insertions and point mutations in the *btk* gene have been identified using Southern blotting, SSCP and sequence analysis (5,15,17). The Btk protein has been shown to be a tyrosine kinase (8), however, we have shown here that it is unlikely to be of major importance in malignant B-cell transformation and leukemogenesis, at least in the B-cell malignancies of childhood. In XLA patients, the block in B-cell differentiation appears to occur after the pre-B cell stage, as the numbers of these cells in their bone marrow are normal. Our findings do not exclude the possibility of *btk* involvement in other leukemias including those where the target cells for malignant transformation are more similar to those absent in XLA patients. It may be of interest, therefore, to characterize *btk* expression in chronic lymphocytic leukemia (CLL), in which B cells of intermediate maturity have been transformed, and also the folicular lymphomas which represent the major group of B-cell malignancies in adults. In conclusion, the fact that the *btk* gene appears to have little, or no, oncogenic potential in the leukemias studied, suggests that it may be useful to develop the techniques for somatic gene therapy to correct the defect in children with X-linked agammaglobulinemia.

*Acknowledgements. We would like to thank Dr Paul Baird for help and advice in setting up the SSCP analysis, the Wellcome Trust, the Leukaemia Research Fund and the Child Health Research Appeal Trust of the Institute of Child Health for financial support.*

## REFERENCES

1. Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia. Leukemia **1988**;2:120–125.
2. Greaves MF, Alexander F. An infectious etiology for common acute lymphoblastic leukemia in childhood. Leukemia **1993**;7:349–360.
3. Bruton OC. Aggamaglobulinemia. Pediatrics **1952**;9:722–727.
4. Rosen FS, Cooper MD, Wedgewood RJP. The primary immunodeficiencies. New Engl J Med **1984**;311:235–242.
5. Vetrie D, Vorechovsky I, Sideras P, Holland J, Davies A, Flintner F, Hammarstrom L, Kinnon C, Levinsky R, Bobrow M, Smith CEI, Bentley D. The gene involved in X-linked agammaglobulinemia (XLA) is a member of the src family of protein-tyrosine kinases. Nature **1993**;361:226–233.
6. Lovering R, Middleton-Price HR, O'Reilly M-AJ, Genet SA, Parkar M, Sweatman AK, Bradley LD, Alterman LA, Malcolm S, Morgan G, Levinsky RJ, Kinnon C. Genetic linkage analysis identifies new proximal and distal flanking markers for the X-linked agammaglobulinemia gene locus, refining its localization in Xq22. Hum Mol Genet **1993**;2:139–141.
7. Parolini O, Hejtmancik JF, Allen RC, Belmont JW, Lassiter GL, Henry MJ, Barker DF, Conley ME. Linkage analysis and physical mapping near the gene for X-linked aggamaglobulinemia. Genomics **1993**;15:342–349.
8. Tsukuda S, Saffran DC, Rawlings DJ, Parolini O, Cutler Allen R, Klisak I, Sparkes RS, Kubagawa H, Mohandas T, Quan S, Belmont JW, Cooper MD, Conley ME, Witte ON. Deficient expression of a B-cell cytoplasmic tyrosine kinase in X-linked aggammaglobulinemia. Cell **1993**;72:279–290.
9. Mano H, Ishikawa F, Nishida J, Hisamura H, Fumimaro T. A novel protein-tyrosine kinase, tec, is preferentially expressed in liver. Oncogene **1990**;5:1781–1786.
10. Siciliano JD, Morrow TA, Desiderio SV. itk, a T-cell specific tyrosine kinase gene inducible by interleukin 2. Proc Natl Acad Sci USA **1992**;89:11194–11198.
11. Kusrock R, Gutterman JU, Talpaz M. The molecular genetics of Philadelphia chromosome-positive leukemias. New Engl J Med **1988**;319:990–998.
12. Nichols J, Nimer SD. Transcription factors, translocations and leukemia. Blood **1992**;80:2953–2963.
13. Sandberg AA. The chromosomes in human cancer and leukemia. New York, Amsterdam, Oxford: Elsevier, **1990**;352–357.
14. Smith DB, Johnston KS. Single step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-s-transferase. Gene **1988**;15:31–40.
15. Bradley LAD, Sweatman AK, Lovering R, Jones A, Morgan G, Levinsky R, Kinnon C. Mutation detection in the X-linked aggamaglobulinemia gene, *btk*, using single strand conformation polymorphism analysis (in press).
16. Chessells JM. Acute lymphoblastic leukaemia. In: Whittaker J, Delamore I, eds. Leukaemia. Oxford: Basil Blackwell, **1987**:331–358.
17. Lovering R, Sweatman A, Genet S, Middleton-Price H, Vetrie D, Vorechovsky I, Bentley D, Fontan G, Espanol T, Morgan G, Levinsky R, Kinnon C. Identifiction of deletions in the *btk* gene allows unambiguous assessment of carrier status in families with X-linked aggamaglobulinaemia (in press).