SAPIENZA
Università di Roma
Facoltà di Scienze Matematiche Fisiche e Naturali

DOTTORATO DI RICERCA
IN GENETICA E BIOLOGIA MOLECOLARE

XXIII Ciclo
(A.A. 2009/2010)

Common fragile sites: structure, replication and their role in chromosome instability

Dottorando
Nazario Bosco

Docente guida
Prof. Franca Pelliccia

Tutor
Prof. Francesca Degrassi

Coordinatore
Prof. Irene Bozzoni

TABLE OF CONTENT

ABSTRACT

Common fragile sites (CFS) are specific regions of mammalian chromosomes that are particularly prone to gaps/breaks. The analysis of these sequences has not definitively clarified the causes of their fragility. There is evidence that CFSs are regions of late/slowed replication in the presence of sequence elements that have the propensity to form secondary structures, and that the cytogenetic expression of CFSs may be due to unreplicated DNA. They are a cause of genome instability, and the location of many CFS correlates with many breakpoints of aberrations recurrent in some cancers.

In the first part of this work the DNA sequence of the CFS FRA7B was determined, and then analyzed to identify structural features potentially involved in fragility. FRA7B maps at the 7p chromosome terminal region and is one of the largest CFS analyzed so far. A 90kb-long sequence that presents very high DNA helix flexibility values was identified. This sequence could have a role in inducing stalling of the replication fork and fragility at FRA7B (Bosco et al., 2010).

In order to analyze the relationship between DNA replication time and fragility, I have also investigated the replication timing of sequences mapping within two CFSs (FRA1H and FRA2G). Our results indicate that the fragile sequences are slow replicating. Thus these regions could sometimes reach mitosis unreplicated and be expressed as chromosome gaps/breakages (Pelliccia et al., 2008).

The last part focused on the role of common fragile sites in the generation of genome instability seen in cancer. Common

fragile sites FRA1H, FRA2G and the later characterized FRA7B, are known to be frequent breakpoints in chromosome aberrations in various neoplasms. A panel of 21 human cancer cell lines was analyzed for the presence of deletions or genome amplification mediated by these CFSs. The results revealed the presence of a duplicated region on a chromosome der(2) present in the karyotype of two analyzed leukemia cell lines K562, K562-1 and K562-2. The two duplicated regions have different lengths and are organized into two large palindromes, which suggests that one BFB cycle has occurred. The localization of the breakpoints were molecularly defined, and are localized in the sequence of three common fragile sites: FRA2H, which was molecularly characterized in this work, the newly identified aphidicolin inducible fragile site FRA2S, and FRA2G (Pelliccia et al., 2010).

INTRODUCTION

1. Fragile sites

The first description of non-random human chromosome fragility was reported in 1965 in cells from a woman previously irradiated (Dekaban, 1965).

The term '**fragile site**' was later introduced by Hecht to describe recurrent chromosome breaks on the long arm of chromosome 16, which segregated in a Mendelian fashion in a large family and showed linkage to the haptoglobin locus (Magenis et al., 1970).

Since that time, fragile sites continued as an active area of research in cytogenetics and their definition and classification were subject to several controversies with the increasing of knowledge.

Today fragile sites are defined as specific loci that preferentially exhibit gaps and breaks on metaphase chromosomes following partial inhibition of DNA synthesis.

Fragile sites are normally stable in cultured cells, but form visible gaps and breaks under certain conditions or treatment with specific chemical compounds (Fig. 1).

To date, according to the Genome Database and to a recent global genome screening, more than 200 different fragile sites have been identified in the human genome (Mrasek et al., 2010).

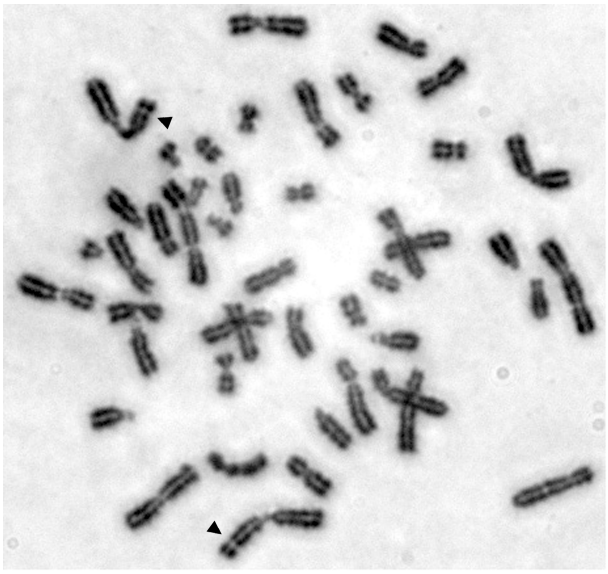


Figure 1. Giemsa stained metaphase obtained from human lymphocyte. The arrowhead shows fragile sites expression.

To date, a chromosome specific nomenclature has been proposed for fragile sites. Each fragile site name starts with the abbreviation FRA followed by the chromosome number and a capital letter according to the appearance of the fragile site description from the short to the long arm of the chromosome, starting from A to Z, e.g. FRA1A was the first fragile site described on chromosome 1 in 1p36.

The study of fragile sites has had a major impact on human genetics, most notably in leading to the identification of the fragile X syndrome and trinucleotide repeat expansion as a mutational mechanism in human genetic disease. Fragile sites have also been associated with genome instability in cancer cells

and with activation of DNA damage response to stalled replication.

Fragile sites are generally categorized into two main classes based on their population frequency and pattern of inheritance. Each class is further subdivided according to their specific mode of induction *in vitro*, i.e. their culture requirements (Sutherland and Hecht, 1985; Tab. 1).

'Rare fragile sites' are present in less than 5% of the population, segregate in a Mendelian manner (Sutherland, 2003), and increased breakage at these sites is most often caused by expansion of nucleotide repeats. The major group of rare fragile sites is the folate-sensitive group associated with CGG-repeat expansion. This group includes FRAXA, in the *FMR1* gene, which is responsible for the fragile X syndrome, and FRAXE in the *FMR2* gene, which is associated with non specific mental retardation (Yu et al., 1991; Gu et al., 1996). Other nonfolate-sensitive rare fragile sites, characterized by expanded AT-rich minisatellite repeats, are induced by bromodeoxyuridine (BrdU) or distamycin A. These include FRA10B and FRA16B, in which alleles with greatly expanded 42- and 33-AT minisatellite repeats are expressed as fragile sites (Hewett et al., 1998; Yu et al., 1997, Tab.1).

'Common fragile sites', which are seen in all individuals, are the largest class of fragile sites. Unlike rare fragile sites, common fragile sites represent a component of normal chromosome structure and are not the result of mutated nucleotide repeat expansion. The cytogenetic expression of these sites varies from individual to individual, reaching a level of expression up to 30% in some (Sutherland and Richards, 1995). Gaps and breaks are visible over wide chromosomal regions of megababses in size. Common fragile sites seem therefore to represent regions of fragility, rather then specific loci as seen for rare fragile sites in which the breakage corresponds to the expanded locus (Handt et al., 2000). Their study has its roots in the early investigations of the fragile X syndrome when

recurrent, site-specific chromosome breaks were noted in cells from both normal and affected individuals when grown under conditions of folate stress used to induce the cytogenetic expression of the fragile X site FRAXA. It was later determined that the great majority of common fragile sites are also specifically and reproducibly induced by low doses of aphidicolin, an inhibitor of DNA polymerases (Glover et al., 1984). Sites of recurrent chromosome breakage on normal chromosomes following treatment of cells with BrdU or 5-azacytidine have also been described and considered as common fragile sites (Sutherland et al., 1985; Tab. 1).

An interesting feature of common fragile sites is their evolutionary conservation among all mammalian species examined to date, e.g. primates, horse, cow, pig, dog, cat, rat, deer, mouse (Ruiz-Herrera et al., 2006, and references cited therein). The fact that common fragile sites persist in widespread phyla suggests they serve for a conserved, to date not understood, important biological role.

Thus, the understanding of the mechanism and the molecular basis of fragile sites is highly important for the comprehension and management of a series of biological processes and human pathologies.

Tab. 1. Classification and induction requirement for fragile sites expression

Class	Mechanism of induction	Sequence
Rare Fragile Sites		
Folate-Sensitive (Folic Acid Deprivation)	Replication stress due to a limiting concentration of either dTTP or dCTP	$(CGG)_n$
Distamycin A	Oligopeptidic antibiotic that binds the DNA in the minor groove impeding DNA replication	AT repeats
BrdU	Base analog of thymidine that delays DNA replication after incorporating itself into the DNA	AT repeats
Common Fragile Sites		
Aphidicolin	Specific inhibitor of the replicative DNA polymerase α and δ	AT-rich
BrdU	Base analog of thymidine that delays DNA replication after incorporating itself into the DNA	AT-rich
5-Azacytidine	Base analogue of cytosine that is incorporated into the DNA sequence affecting DNA replication/condensation	AT-rich

2. Common fragile sites

In cultured cells, common fragile sites are hotspots for metaphase chromosome gaps/breaks and induced chromosomes rearrangements. After induction, common fragile sites are regions of potential genome instability (Glover, 1998), being hotspots for deletion and translocations (Glover and Stein, 1988; Wang et al., 1997), increased rates of sister chromatid exchange (Glover and Stein, 1987; Feichtinger and Schmid, 1989; Hirsch, 1991), plasmid or viral integration (Rassool FV et al., 1991; Wilke et al., 1996; Mishmar et al., 1998) and intrachromosomal gene amplification (Coquelle et al., 1997; Hellman et al., 2002; Debatisse et al., 1998). Based on these characteristics and the association with cancer breakpoints on banded chromosomes, a number of early reports suggested that common fragile sites could be responsible for some of the chromosome rearrangements observed in cancer (Hecht and Glover, 1984; Yunis and Soreng, 1984).

At present more then twenty common fragile sites have been cloned and characterized at the molecular level in various ways (Mrasek et al., 2010). The genomic breakage and instability at these sites occur along a large genomic region extending over at least 500 kb (Paradee et al., 1996). All are relatively AT-rich areas (Boldog et al., 1997), but they do not show any repeat motifs such as expanded trinucleotide or minisatellite repeats that could predispose to fragility as has been demonstrated in the rare fragile sites (Arlt et al., 2002; Mishmar et al., 1998; Paradee et al., 1994). To understand the molecular mechanism of fragility at these sites, another investigative approach was adopted in which structural characteristics of the DNA rather than their sequence *per se* was examined.

When measuring local fluctuations at the twist angle between consecutive base pairs along the DNA molecule backbone, it was found that many of the common fragile site

regions analyzed to date, contain more areas of high DNA torsional flexibility, termed flexibility peaks, than non-fragile regions mapped to the same chromosome band (Arlt et al., 2002; Schwartz et al., 2006; Fig. 2A). These flexibility peaks, which can be present as 'cluster of flexibility peaks' when including at least three close-set peaks (Zlotorynski et al., 2003), are composed of interrupted AT-dinucleotide-rich sequences of various length termed AT-dinucleotide-rich flexibility islands. These islands, which are significantly more AT-dinucleotide-rich than their nonflexible flanking sequences, have the potential to form unusual DNA secondary structures that can perturb replication (Zlotorynski et al., 2003; Fig. 2).

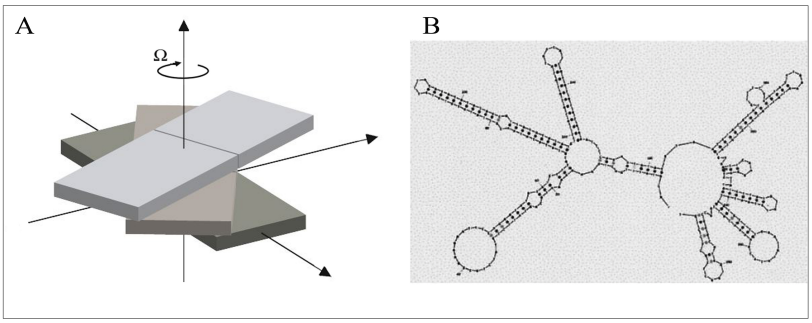


Figure 2. A. Schematic representation of the twist angle of DNA helix. B. Example of predicted secondary structure by Mfold program of an AT-dinucleotide rich flexibility island (taken from Zlotorynski et al., 2003).

The AT-rich flexible regions contained in the common fragile sites are similar to the AT-minisatellite repeat of the FRA16B and FRA10B rare fragile sites. These sequences have the potential to form secondary structures able to affect DNA replication (Zlotorynski et al., 2003). Moreover, the FRA16B AT minisatellite exclude nucleosomes affecting the chromatin

structure (Hsu and Wang, 2002). Thus, the molecular basis of rare and common fragile sites may actually be the same or very similar. However the molecular mechanism underlying the genetic instability at fragile site remains to be investigated.

3. Replication pattern at common fragile sites

The two conditions that induce expression of the majority of human fragile sites, folate deprivation and aphidicolin, both inhibit DNA replication, leading to the idea that either stalled or collapsed replication forks or unreplicated DNA that persists into metaphases is the cause of fragile site expression.

Supporting the idea that fork stalling is important, 2D gel analysis of expanded CGG/GCC repeat sequences analyzed in both *E. coli* and yeast cells directly demonstrated stalled or slowed replication forks that mapped to the repeated sequence (Samadashwily et al., 1997; Pelletier et al., 2003). Recent data from Freudenreich's lab, have shown that a short sequence that includes the high flexibility peaks within the common fragile site FRA16D, stalls a replication fork when replicated on a plasmid in yeast cells (Zhang and Freudenreich, 2007). The AT repeats embedded in the FRA16D flexible region are predicted to easily form either an hairpin (from a single-strand of DNA) or cruciform (from double-stranded DNA) structure (Zhang and Freudenreich, 2007). In addition the flexible sequence significantly increases chromosome breakage, which is further enhanced by either the presence of replication inhibitors or absence of the DNA repair protein Rad52.

Replication timing of fragile sites was first evaluated at the rare FRAXA site (Hansen et al., 1993). It was found that the normal allele of the FRAXA region replicates in late S phase, while replication of alleles with CGG expansions was delayed until G2/M. Interestingly, the region of replication delay was

quite large, 400 kb or more on either side of the repeat expansion, suggesting that more than a single stalled fork is involved (Subramanian et al., 1996). The AT minisatellite expansion at FRA10B and FRA16B also result in a replication delay, although the region of delay for FRA10B mapped somewhat distal to the repeat expansion (Handt et al., 2000).

Common fragile sites FRA3B and FRA16D are also both late replicating, and for FRA3B exposure to aphidicolin delays replication further, resulting in a failure to complete replication in some cells (Le Beau et al., 1998; Palakodeti et al., 2004). FRA3B replication timing studies in different cell lines showed that the replication of this site is asynchronous, one allele replicate later than the other, with or without aphidicolin treatment, and that fragility was preferentially observed on the late replicating allele (Wang et al., 1999). This result suggests a possible relationship between particular alleles and fragility. Common fragile site FRA7H has a more complicated pattern of replication with allelic asynchrony that is further enhanced by aphidicolin (Hellman et al., 2000). This latter pattern is consistent with a stochastic inhibition or slowing of replication, such as would happen at a sequence with a potential to stall a fork.

Furthermore, mammalian chromosomes present two distinct genome domains: G-bands and R-bands. G-bands are regions with high A/T content, replicate late in S-phase, are relatively insensitive to DNase I, are gene poor and rich in LINE elements. Complementary R-bands, instead, are G/C rich, replicate early in S-phase, are DNase sensitive, are rich in gene and Alu elements (Gardiner, 1995). Takebayashi et al. (2001) showed that replication forks typically slow down during the early-mid S-phase, when R/G transition occurs. In a later work from the same lab, was shown that the replication timing switches precisely from early to late S phase, when the replication forks proceed through the putative R/G chromosomal band boundary predicted by marked changes in GC content, at a

sequence level, of the region 1q31.3 (R-band)-1q32.1 (G-band) (Takebayashi et al., 2005). These data suggest that the slowdown of replication is a general feature of R/G band boundaries existing throughout the genome. Common fragile sites usually map at band boundaries. Moreover, some common fragile sites mapping to R-bands share structural features with the chromatin of G-bands, simulating a G/R band boundary (Mishmar et al., 1999). This different chromatin organization of common fragile sites might affect the replication and condensation of the fragile sequences and thus contribute to the fragility.

There could also be a link between replication origins and fragile sites, as peaks of high flexibility are over-represented at mapped mammalian replication origins. Furthermore, FISH analysis of aphidicolin-treated hamster cells revealed fragile site expression at two origins (Toledo et al., 2000). The same group has also shown that nucleotide pool levels can modify origin usage, so that a pattern of one dominant origin can switch to a pattern of many weaker origins during condition of replication stress (Anglana et al., 2003). Thus, either a change in origin usage or more persistent origin bubbles could create areas of increased susceptibility to breakage. For example, one might imagine that if a previously inactive high flexibility sequence became an active origin in the presence of aphidicolin, it could give opportunity for formation of a secondary structure that would preclude normal replication and become a fragile site (Freudenreich, 2007).

Furthermore, in general AT islands are hypervariable elements due to polymerase slippage during replication and/or unequal recombination events (Bois and Jeffreys, 1999). The flexibility, thermodynamic instability and propensity to form superhelical duplexes of AT-rich regions are consistent with the idea that AT islands may serve as anchorage sites for DNA on the nuclear matrix (Matrix Attachment Regions, MARs) (Woynarowski et al., 2001). It’s known that MARs are regions

prone to unwinding and bending, contain topoisomerase II binding and cleavage sites and possibly can acts as DNA replication origins. Thus these regions may constitute hot spots for breakage and DNA recombination and may be involved in chromosome fragility (Svetlova et al., 2001).

4. Common fragile site and cancer

Numerous studies have shown that common fragile sites are frequently involved in chromosome breakages and rearrangements in cancer cells. The fragile site-specific rearrangement most frequently observed is one or more large deletions of ten to hundreds of kilobases directly within the fragile region, resulting in inactivation of associated genes.

Most studies have focused on FRA3B and FRA16D, because they are the two most frequently expressed and best-characterized common fragile sites, and both lie within the large tumor suppressor genes, *FHIT* and *WWOX*, respectively.

The *FHIT* (fragile histidine triad) gene catalyzes hydrolysis of diadenosine polyphosphates, produced via action of the aminoacyl-tRNA synthetases, and is the major diadenosine triphosphate (ApppA) hydrolase found in mammals (Barnes et al., 1996). Its function has been linked to intracellular signaling and the DNA damage response (Pekarsky et al., 2004; Shi et al., 2000). There is ample evidence that FHITp acts as a tumor suppressor. *Fhit*-deficient mice have increased susceptibility to N-nitrosomethylbenzylamine (NMBA)-induced gastric tumors, which can be rescued by introduction of a functional *Fhit* allele (Zanesi et al., 2001). In addition, overexpression of *FHIT* suppresses growth of different cancer cell lines both *in vitro* and *in vivo* (Siprashvili et al., 1997). *FHIT* is frequently involved in biallelic loss and other chromosome abnormalities in tumors (Huebner and Croce, 2003; Ohta et al., 1996). *FHIT* deletions, abnormal transcripts,

promoter hypermethylation, and associated loss of expression are common in human malignancies. In cancer cells, these events are often associated with deletions directly within the FRA3B region, centering on exon 5 of *FHIT*.

Recent results from Aqeilan et al. (2007) and others (Bednarek et al., 2001; Paige et al., 2001) also demonstrate a tumor suppressor function for the *WWOX* gene. *WWOX* encodes a 46-kDa protein that contains two WW domains, a compact 38 amino acid residue units that fold into a three-stranded β-sheet structure, and a short-chain dehydrogenase/reductase domain that may function in apoptotic pathways (Chang, 2002; Chang et al., 2001, 2003). Deletions of microsatellite markers within the FRA16D/*WWOX* locus are prevalent in several tumor types including breast, prostate, esophageal, lung, stomach, and pancreatic carcinomas (Durkin and Glover, 2007).

Furthermore, similar deletion patterns in cancer cells have also been shown for other common fragile sites and associated genes (Tab. 2). Thus, a considerable amount of data show that many common fragile sites represent targets that are 'weak links' for genomic alterations in at least some cancer cells (Huebner and Croce, 2001).

Furthermore, recent findings suggest that both common and rare fragile sites contain a number of microRNA (Calin et al, 2004). Over half of the mapped microRNA lies in known fragile chromosome regions. The high level of evolutionary conservation observed at common fragile sites might be the result of important regulatory sequences found there that include miRNA genes, which can be encoded in both intronic and intergenic regions. Interestingly, many characterized common fragile sites lie inside very large genes.

While deletion breakpoints within common fragile sites in cancer cells are common, relatively fewer translocations involving fragile sites have been reported. Translocation involving FRA3B have been found in a small number of tumor cell lines, including hepatocellular, esophageal, and breast

carcinoma (Fang et al., 2001; Keck et al., 1999; Popovici et al., 2002). On the contrary, FRA16D is a site of frequent translocations in multiple myelomas. As many as 25% of all multiple myelomas contain a recurrent translocation between 16q23 and 14q32, resulting in at least one truncated allele of *WWOX* in these tumor cells (Chesi et al., 1998). Translocations have also been reported at several others common fragile sites, including FRA6E and FRA6F in acute lymphoblastic and acute myeloid leukemia, and FRA2G in members of a family with multifocal clear renal cell carcinoma (Sinclair et al., 2005). The relatively low numbers of reported translocations versus deletions may simply reflect the processes of biological selection in tumor cells. Both deletions and translocations at common fragile sites appear simply to inactivate associated genes rather than deregulate their expression or create fusion genes with altered functions. As such, there may not be any difference in selective advantage conveyed by these two types of chromosomal rearrangements.

Table 2. Genes and murine orthologs associated with cloned Common Fragile Sites (from Durkin and Glover, 2007).

Human CFS	Location	Associated genes	Murine ortholog
FRA2G	2q31	*IGRP, RDHL, LRP,*	Fra2D
FRA3B	3p14.2	*FHIT*	Fra14A2
FRA4F	4q22	*GRID2*	Fra6C1
FRA6E	6q26	*PARKIN, MAP3K, LPA*	Not identified
FRA6F	6q21	*REV3L, DIF13, FKHRL*	Not identified
FRA7E	7q21.2	*LEP*	Not identified
FRA7G	7q31.2	*CAV1, CAV2, TESTIN, MET*	Fra6A3.1
FRA7H	7q32.3	Not identified	Fra6B1
FRA7I	7q36	*PIP*	Not identified
FRA7K	7q31.1	*IMMP2L*	Fra12C1
FRA8C	8q24.1	*MYC*	Not identified
FRA9E	9q32–33.1	*PAPPA, ROD1, KLF4*	Fra4C2
FRA16D	16q23.3	*WWOX*	Fra8E1
FRAXB	Xp22.3	*STS*	Not identified

In addition to deletion and translocation breakpoints, common fragile sites have been associated to viral integration and gene amplification in tumor cells. The FRA3B region was also found to contain human cervical cancer HPV-16 integration sites (Wilke et al., 1996), which led to findings of additional viral integration sites at common fragile sites in tumors or tumor cell lines (Popescu and DiPaolo, 1989; Smith et al., 1992; Thorland et al., 2000).

Gene amplification, that is the increase of copy number of a definite chromosome region, is a genome alteration observed in many human neoplasms that produce overdosage of oncogenes mapped in the amplified regions. The identification and analysis of these regions therefore often allow the discovery of new putative oncogenes.

The amplified sequences can be present both in the continuity of a chromosome as '**homogeneous staining regions**' (HSR), and as small free chromatin structures known as '**double minutes**' (DM). The mechanisms underlying these manifestations of genome instability are under investigation. In particular, there is some evidence that common fragile sites can initiate **breakage-fusion-bridge (BFB) cycles**, a mechanism responsible for accumulation of intrachromosomal (HSR) amplicons (Coquelle et al., 1997; Kuo et al., 1994; Fig. 3). More recently, FRA7I and FRA7G have been identified at one boundary of the amplicons found in two tumor-derived cell lines (Hellmann et al., 2002).

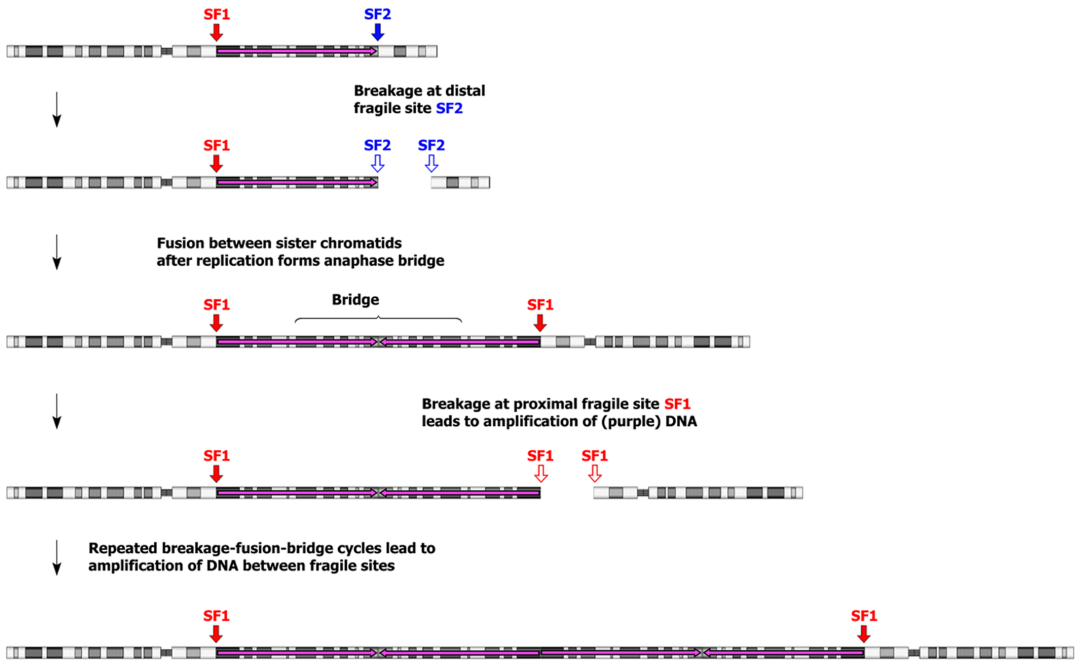
In the model of intrachromosomal gene amplification obtained by break at common fragile sites, the amplicons appear to be arranged in head-to-head organized ladders obtained by BFB events (McClintok, 1951, Fig. 3). Two common fragile sites may set boundaries of the early amplicon: the break at the more telomeric site determining the telomeric loss and then the sister chromatid fusion, and the break at the more centromeric site defining the amplicon size (Coquelle et al., 1997 and references cited therein). BFB cycle repetition is probably

interrupted when the amplified chromosome is stabilized by the addition of a telomere. Cytogenetic analyses have demonstrated the validity of this model in rodents (Coquelle et al., 1997; Kuo et al., 1994), and in human cancer (Shuster et al., 2000; Ciullo et al., 2002; Hellmann et al., 2002; Zimonjic et al., 2003; Reshmi et al., 2007).

5. DNA damage checkpoints regulate common fragile site stability

Little was known of the cellular mechanisms controlling common fragile site stability until it was discovered that the Ataxia-telengiectasia and Rad3 Related (ATR) checkpoint kinase plays a major role in regulating fragile site stability (Casper et al., 2002). During S and G2 phases of the cell cycle, when DNA is replicated and chromosomes prepared for mitosis, the ATM (Ataxia Telengiectasia Mutated) and ATR kinases act in parallel, as a major DNA damage checkpoint proteins in overlapping pathways (Harrison and Haber, 2006). ATM responds primarily to DNA double strand breaks, while ATR orchestrates the signaling of downstream effector molecules that respond to stalled and collapsed replication forks. Cells lacking functional ATR are deficient in checkpoint responses to UV light and agents that block replication fork progression, including hydroxyurea, aphidicolin and hypoxia conditions (Nghiem et al., 2001).
Based on the appearance of chromosome gaps and breaks at common fragile sites following replication stress and the mechanism of action of agents that induce cytogenetic fragile site expression, Casper and colleagues (2002) studied the effects of ATM and ATR deficiency of these proteins on common fragile site breakages. ATR, but not ATM, was found to play a major role in maintaining the stability of common fragile sites, by directing the cellular checkpoint response to stalled replication at these sites.

Figure 3. A model for intrachromosomal genome amplification mediated by fragile sites (SF1 in red and SF2 in blue). Initial breakage at distal fragile site (SF2) leads to sister chromatid fusion after replication. At anaphase, the dicentric chromosome appears as a bridge between centromeres moving to opposite poles of the mitotic spindle. Breakage of this giant inverted repeat (in SF1) leaves each daughter cell with a chromatid lacking one telomere, which again fuses after replication, perpetuating the BFB cycles. Amplification occurs in one daughter cell when the breakage is asymmetric.



Cells lacking ATR, in fact, showed a dramatic increase in fragile site expression following treatment with low doses of aphidicolin. Furthermore, ATR deficiency alone, without addition of replication inhibitors, induced a low frequency of spontaneous gap and breaks at common fragile sites, showing that ATR is required for common fragile site stability even during unperturbed replication, and indicating that some level of replication stalling occurs normally at fragile site regions.

In contrast, ATM-deficient cells did not exhibit spontaneous or aphidicolin-induced common fragile site breaks. This lack of increased breakage in ATM-deficient cells suggest that double strand breaks are not the initial or primary cause of cytogenetic common fragile site expression. However, the ATM pathway may be important in regulating subsequent event at common fragile sites, particularly in the resolution of double strand breaks that must occur in these regions to give rise to chromosomal rearrangements.

Casper et al. (2004) subsequently found that cells from individuals from Seckel syndrome, which contain a hypomorphic mutation in ATR, show dose-responsive increased breakage at common fragile sites.

The discovery that ATR is intimately involved in regulating the stability of common fragile sites was important to linking cell cycle checkpoint function to fragile site stability.

Subsequent investigation have focused on further delineation of these mechanisms, and a number of targets or modifiers of the ATR-regulated pathway have now been shown to influence common fragile site stability, including BRCA1, CHK1, the Fanconi Anemia (FA) pathway proteins, and the SMC1 (Arlt et al., 2004; Durkin et al., 2006; Howlett et al., 2005; Musio et al., 2005).

Howlett et al. (2005) examined the role of the Fanconi Anemia pathway in regulation of common fragile site stability. The FANCD2 protein is activated via mono-ubiquitination during S-phase, signaling its translocation to BRCA1- and

RAD51-positive nuclear foci (Taniguchi and D'Andrea, 2002). It had also been shown that ATR phosphorylates FANCD2 and is required for its efficient mono-ubiquitination (Andreassen et al., 2004). These results suggested a role for the FA pathway in the response to DNA replication stress and so, in the maintenance of common fragile site stability. More recently, Hickson's group showed, that replication inhibitors such as aphidicolin induce sister chromatid DNA bridges as replication intermediates, specifically at common fragile site loci. These structures, marked by FANCD2/FANCI foci, inefficiently resolved before anaphase by DNA repair protein complexes, will lead to anaphase bridges, then to chromosome breakages, and so forth, to all events derived from these breakages (Chan et al., 2009).

It has been suggested that breaks at common fragile sites may serve as a 'signature' of stalled or delayed replication in tumor cells, aggravated by deficiencies in the S-phase and G2/M checkpoints or associated repair genes during tumorigenesis (Casper et al., 2002). Concomitant with these events was a high frequency of LOH at known common fragile regions, including FRA3B. These findings suggest that in precancerous lesion, replication stress leads to stalled or collapsed replication forks, resulting in the activation of the ATR- and, with subsequent DNA double strand breaks, the ATM-dependent checkpoints. Cells that do not undergo apoptosis or cell cycle arrest will develop deletions and allelic imbalances that could preferentially target common fragile sites, since they are most sensitive to replication stress (Durkin and Glover, 2007). Further mutations in p53 or other genes will release additional checkpoint restrictions and lead to tumor progression. These findings suggest that lesions at common fragile sites are indicators of replication stress during early stages of tumorigenesis and might explain why deletions within fragile sites are so frequent in cancer cells.

Given the high frequency of instability at common fragile sites and associated sensitivity to replication stress, the question has often arisen whether fragile site-mediated inactivation of associated genes in cancer cells support a functional role for these genes in cancer progression or simply represent a 'bystander effect' of fragile site instability. As more and more common fragile site are characterized at the molecular level, it appears that many are associated with deletions and rearrangement in different tumor cells. There is ample evidence to support a tumor suppressor function of genes such as *FHIT* and *WWOX*. However, deletions of genes with no obvious role in tumor progression, such as steroid sulfatase (STS), found at FRAXA, have also been observed in tumor cells (Arlt et al., 2002). Thus, both answers to this question may be correct, in that lesions at common fragile sites can serve as 'signature' of replication stress during tumorigenesis, and some associated genes, can confer a selective growth advantage to cells in which they are deleted.

Based on the appearance of deletions and chromosome rearrangements in cultured and tumor cells, double strand breaks can clearly occur at common fragile sites, either directly or as a consequence of the misrepair of stalled replication forks. Homologous Recombination Repair (HRR) plays a major role in responding to double strand breaks and stalled or collapsed replication forks during S and G2, when the sister chromatids are present. It has been hypothesized that SCEs (Sister Chromatid Exchanges) are formed by the action of HRR during replication. Glover and Stein (1987) reported that, on average 70% of all gaps and breaks at FRA3B after aphidicolin treatment had a SCE at that site, suggesting a role for HRR in the repair of lesion at common fragile sites. Furthermore, aphidicolin-induced replication stress leads to RAD51 focus formation and phosphorylated DNA-PKcs, key components of the HRR and NHEJ (Non-Homologous End-Joining) double strand breaks repair pathways, respectively (Schwartz et al., 2005). The

downregulation of RAD51, DNA-PKcs, or LIGIV, an additional component of the NHEJ repair pathway, leads to a significant increase in replication stress-induced common fragile site expression in MCF7 cells. Replication stress also resulted in formation of foci of double strand break markers, MDC1 and γH2AX, which colocalized with those of RAD51 and phospho-DNA-PKcs (Schwartz et al., 2005). These findings support the idea that double strand breaks are formed at common fragile sites as a result of replication perturbation and provide the first clues into how lesions are repaired at common fragile sites. Additional mechanistic studies will undoubtedly provide greater detail of the exact nature of the lesions at common fragile sites and their repair.

6. Mechanism of common fragile site instability: a working model for fragile site expression

The identification of CGG trinucleotide repeat expansion as the molecular basis for the appearance of the rare fragile sites, e.g. FRAXA in the *FMR1* gene, suggested early on that common fragile site breakage might have a related mechanism. However, unlike rare fragile sites, no expanded di- or trinucleotide repeat sequences have been identified within common fragile sites. Nonetheless, sequences, and perhaps repetitive elements, could still be important factors affecting common fragile sites instability.

Common fragile sites contain a relatively high AT-rich sequence including long stretches of perfect AT microsatellite sequence, or AT-islands with high DNA flexibility, at least in some analysed common fragile sites. Zlotorynski et al. (2003) have shown that the sequences at flexibility peaks are composed of interrupted runs of AT-dinucleotides, and these sequences show similarity to the AT-rich minisatellite repeats that underlie the fragility of the rare fragile sites FRA16B and FRA10B. Such

sequences have the potential to form secondary structures and, hence, may affect replication at fragile sites.

If common fragile sites require specific sequences for their 'fragility', then large deletions that remove the necessary sequences should cause the affected common fragile sites to lose their 'fragility'. Arlt et al. (2002) found that two tumor cell lines containing a 500 kb deletion of FRAXB completely eliminated fragility at this site. However on contrary, Corbin et al. (2002) examined FRA3B in hybrid cells containing tumor-derived chromosomes 3 with large FRA3B intralocus deletions and found that the deletions did not reduce aphidicolin-induced fragility. In a recent intensive study of somatic cell hybrids containing large 200-600 kb deletions of FRA3B that deleted several major flexibility peaks, fragility of this site was significantly reduced but not absent (Durkin and Glover, 2007). These results suggest that loss of large blocks of sequences within common fragile sites reduces, but does not result in complete loss of fragility and that common fragile site expression on metaphase chromosomes likely reflects the effects of a number of sequences across the large fragile region.

Ragland et al. (2008) showed that integration of sequences from FRA3B are sufficient to recapitulate fragile site-like instability at a novel genomic location. Cell clones containing stable integrated FRA3B sequences at unique nonfragile site loci retained fragile site-like instability at ectopic sites. However they found that the fragility of integrated common fragile site sequences was not dependent on late replication.

These data suggest that for common fragile sites there must not be only a single feature that makes these sequences 'fragile', but further analyses are required to study in depth the mechanisms that make fragile sites 'fragile'.

Summarizing what has been said so far, based on aphidicolin-mediated inhibition of DNA polymerase, Glover et al. (1984) proposed, in the initial descriptions of common fragile

sites, that breakage was caused by preferential inhibition of DNA replication at these sites. Since then, numerous lines of evidence continue to support this basic hypothesis. The underlying molecular basis for stalled replication following replication stress is not entirely understood but likely relates to sequence and late replication at fragile sites. To date, data from a number of laboratories support a working model for common fragile sites expression. This model is based on late or delayed replication, unusual sequence composition, and the role of checkpoint and repair proteins in fragile site stability (Fig. 4). It suggests that the AT-rich sequences at common fragile sites, present difficulties during replication that are further exacerbated by aphidicolin and certain other forms of replication stress. When cells are treated with low dose of aphidicolin, the polymerase slow or pause, likely leaving the helicase/topoI complex to continue unwinding DNA ahead of it, and resulting in long stretches of ssDNA that can activate the ATR-dependent DNA damage checkpoint. These ssDNA regions may form secondary structures, such as hairpins or cruciforms at the AT-repeats, which can further perturb replication as the polymerases encounter them. Although the majority of these perturbations are likely detected by ATR damage checkpoint and DNA repair machinery, some escape and present themselves as gaps and breaks on metaphase chromosomes, particularly at common fragile sites.

This model is also supported by the observation that treatment of cells with low doses of camptothecin, an inhibitor of topoisomerase I, can almost completely prevent the breaks at common fragile sites induced by aphidicolin (Arlt and Glover, 2010). These results suggest that polymerase-helicase uncoupling may be an initial key event in common fragile site instability after replication perturbation and provide new insight into early conditions required for fragile site breakage.

Recently Le Beau has proposed that a unique epigenetic pattern may underlie the unusual sensitivity of common fragile

sites to replication interference (Jiang et al., 2009). They analyzed chromatin modification patterns within the six human common fragile sites with the highest levels of breakage, and their surrounding non fragile regions. Chromatin at most of the common fragile sites analyzed has significantly less histone acetylation than the control non fragile regions. Moreover, Trichostatin A and/or 5-azadeoxycytidine treatments reduced chromosome breakage at common fragile sites. These data show that histone hypoacetylation is a characteristic epigenetic pattern of common fragile sites, and suggest that chromatin at common fragile sites might be relatively more compact, underlying a role for chromatin conformation in fragile site expression. Furthermore, lack of histone acetylation at common fragile sites may contribute to the defective response to replication stress observed in these regions.

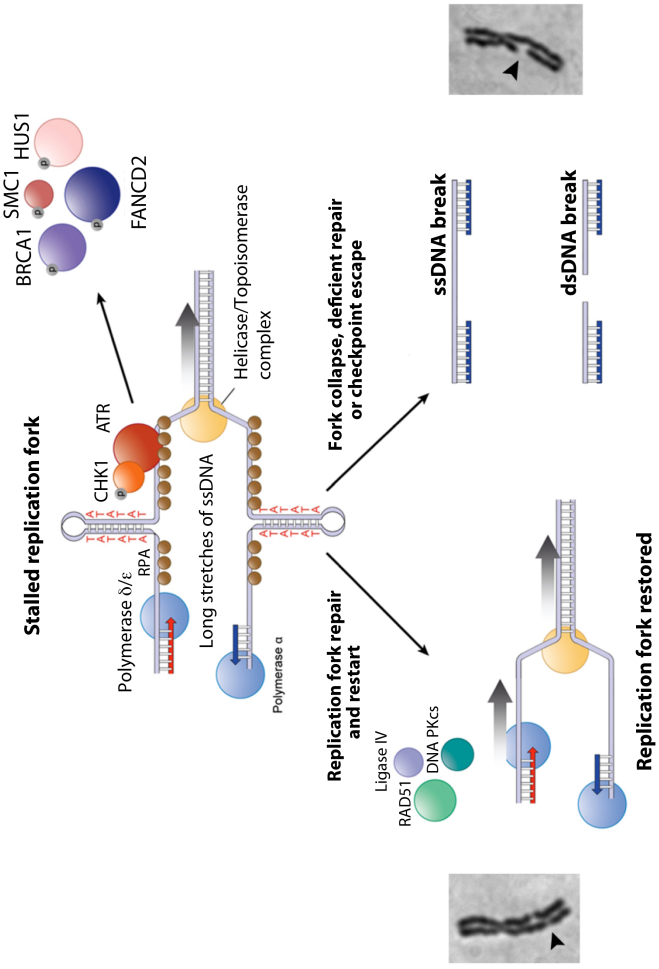


Figure 4. A model for common fragile site expression. This model predicts that fragile sites are derived from long stretches of unreplicated ssDNA that are exposed when a replication fork is stalled or delayed, for example by treatment with APH. Repair of these regions restores replication fork progression. However, sometimes these regions escape checkpoint activation or are left unrepaired, resulting in an unreplicated region that can appear as a fragile site on metaphase chromosomes or lead to a DSB (modified from Durkin and Glover, 2007).

DAPI INDUCIBLE FRAGILE SITES

DAPI (4'-6-diamidino-2-phenilindole) is a non-intercalating fluorescent compound that binds specifically to AT bases of double-stranded DNA in the minor groove (Fig. 5). Footprinting experiments show that a sequence of $(AT)_{n>4}$ is covered when DAPI is bound (Jeppesen and Nielsen, 1989; Portugal and Waring, 1988). When DAPI molecules are bound contiguously to AT-rich areas, the DNA conformation is changed, favoring further compound binding.

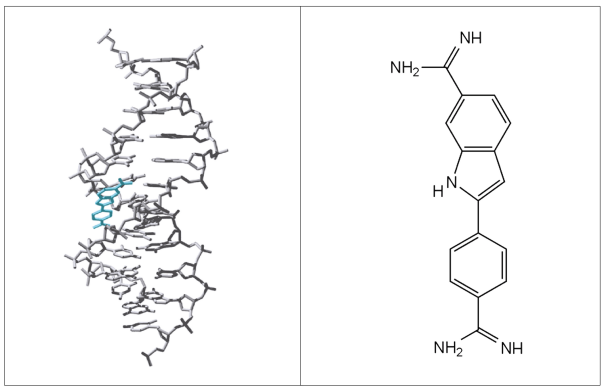


Figure 5. A model for DAPI (4'-6-Diamidino-2-Phenilindole) DNA binding in minor groove, and its molecular structure.

On human chromosomes, DAPI inhibits the complete spiralization of some well-defined chromosomal areas (Matsukuma and Utakoji, 1978). DAPI can also act as an undercondensing agent when supplied to cells after DNA synthesis, during G2 phase (Prantera et al., 1981). When human lymphocytes are grown in complete media, this compound induces a sub-normal condensation of the heterochromatic areas of chromosomes 1, 9, 16 and Yq, as well as four gaps and breaks on chromosomes 1q, 2q, 7p and 13q (Pelliccia and Rocchi, 1986). After chromosome banding, these gaps/breaks were mapped at specific chromosomal bands and have been identified as common fragile sites induced by different treatments: FRA1H, at 1q41-42, induced by 5-azacytidine (Sutherland et al., 1985); and the three aphidicolin inducible common fragile sites FRA2G, at 2q24-31, FRA7B at the terminal region of the short arm of chromosome 7, FRA13C, at 13q21-22 (Glover et al., 1984; Yunis and Soreng, 1984). This group of fragile sites could be considered as a new DAPI-inducible class (Fig. 6).

.

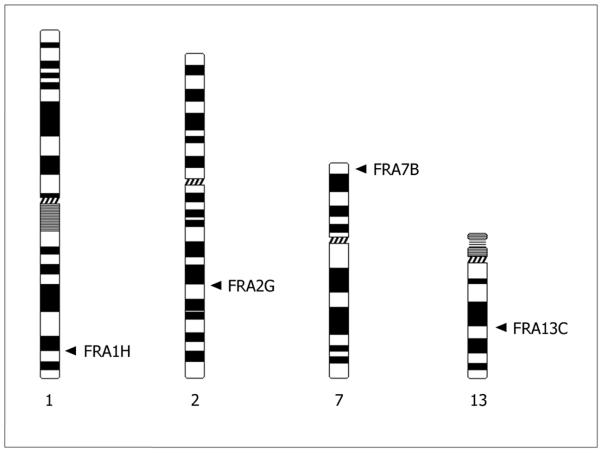


Figure 6. The DAPI-inducible common fragile site family.

Rocchi and Pelliccia (1988) suggest that the large heterochromatic areas and the four common fragile sites induced by DAPI consist of late replicating areas that possess DAPI-specific binding sequences. The combination of late replication (as an intrinsic characteristic of common fragile sites) with the presence of sequences receptive to DAPI, make this uncondensed region available for further DAPI binding and thus no longer capable of undergoing normal condensation. This hypothesis is in agreement with what happens in the heterochromatic areas, in which the frequency of subcondensation induced by DAPI is very high. These areas are, in fact, very AT-rich and late replicating. In conditions that further delay their replication (such as thymidylate stress, e.g. growth in 199 medium) their sensitivity to DAPI is highly increased.

FRA1H, spanning about 10 Mb, maps at the distal part of the G-band 1q41 and at the proximal part of the R-band 1q42.1, thus lying on the boundary of bands with different replication timing in S-phase (Curatolo et al., 2007, Fig. 7). FRA1H is the first characterized common fragile site whose expression is not induced by aphidicolin but instead by DAPI and 5-azacytidine. The analysis of the fragile region with the FlexStab program showed the presence of an AT-rich 120 kb long region that has a very large number of flexibility peaks. This sequence is located at the beginning of the most fragile FRA1H region and, because of its potential to form secondary structures and so to disturb the replication, could be involved in inducing fragility in the surrounding regions (Curatolo et al., 2007). Nine of the genes localized within the FRA1H region, because of their putative role in tumor suppression function, were investigated for homozygous deletion in a panel of heterogeneous cancer cell lines (Pelliccia et al., 2007). The same nine genes were checked for loss or modification of their expression in the same panel of cancer derived cell lines. Five of the analyzed genes showed a significant modification in the

expression level. Interestingly a cluster of two microRNA, *MIRN194-1* and *MIRN215*, showed a detectable expression of the polycistronic RNA in different leukemia and lymphoma derived cell line that is absent in normal lymphocyte (Pelliccia et al., 2007). This finding suggested that the five genes, because of their functions and the modifications of the level of expression shown in some of the analyzed cancer cell lines, could have a role in the neoplastic transformation of some cell type.

Limongi et al. (2003) showed that the most fragile region of the **FRA2G** common fragile site extends over 1 Mb, at the 2q31 region (Fig. 8). An analysis of DNA sequence spanning the fragile region showed a high frequency of flexibility regions within this site that could contribute to fragility. Furthermore, the Burkitt lymphoma-derived cell line DAUDI has a biallelic deletion that involves eight of nine genes chosen for a transcriptional analysis of the FRA2G genes (Fig. 8). Moreover, loss of expression of the *DHSR9* gene was observed in two lymphocyte derived cell lines, MOLT-14 and Raji.

FRA7B and **FRA13C** are not yet characterized. Their characterization would be important to continue the analysis of this particular class of common fragile sites.

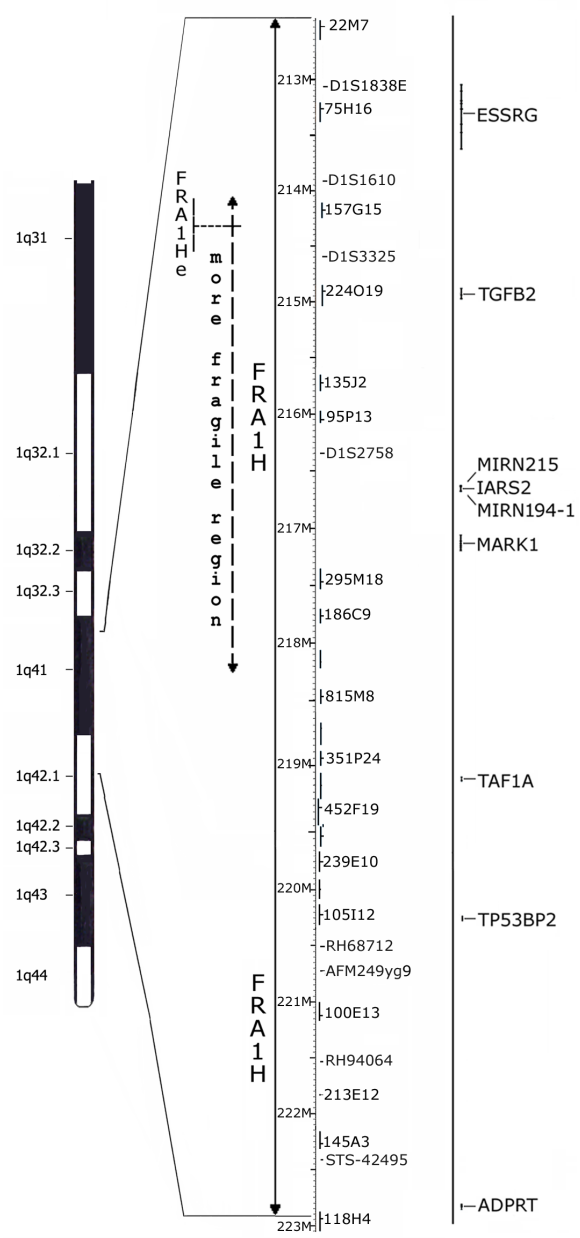


Figure 7. Physical map of the FRA1H region. The BAC and PAC clones used as probes for its characterization, some markers and some genes are shown (Curatolo et al., 2007)

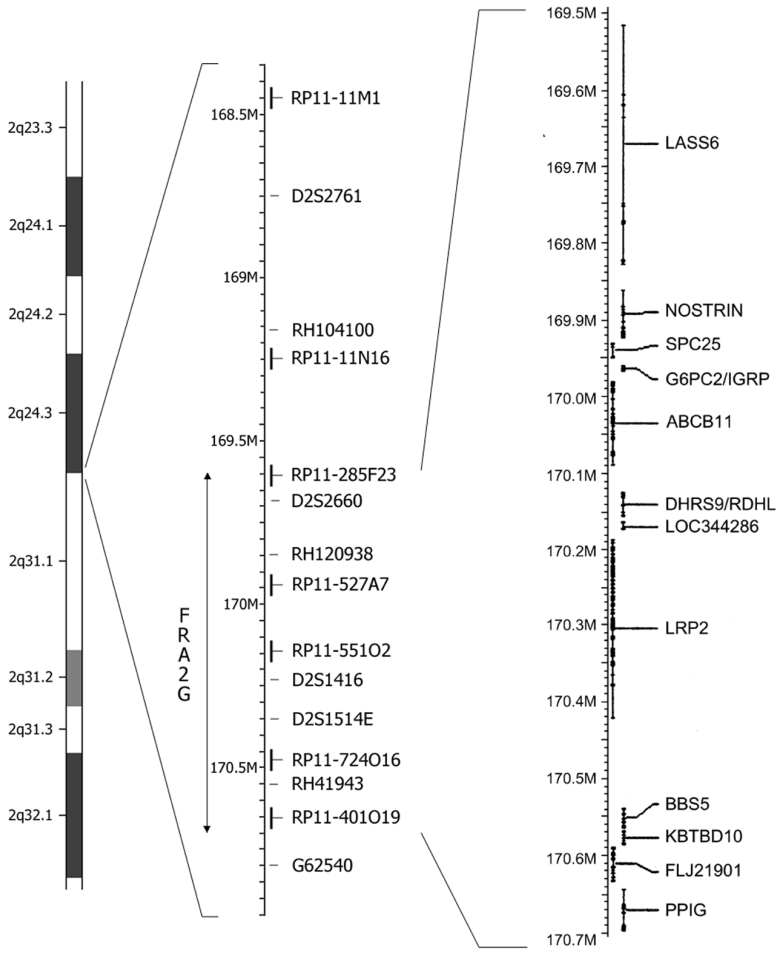


Figure 8. Physical map of the FRA2G region. The BAC clones used as probes for its characterization, some markers and some genes are shown (modified from Limongi et al., 2003).

AIM OF THE RESEARCH

Understanding the molecular basis of common fragile site instability is important given that these regions are frequently deleted or rearranged in cancer cells. Furthermore, common fragile sites provide a unique window into the molecular events that occur following certain types of replication stress in mammalian cells. Due largely to their instability in tumor cells and the link to important cell cycle checkpoint and DNA repair pathways, the study of common fragile sites has become increasingly important over the past few years, and a great deal has been learned about their genomic structure and mechanisms of instability. However, a number of questions remain, and future studies should enhance our understanding of the dynamics of common fragile sites replication and the significance of these conserved regions in normal and cancer cells.

Due to the lack of knowledge about chromosomal fragility, and to further investigate the DAPI inducible class, in the first part of my work, I have molecularly characterized the common fragile site FRA7B (Bosco et al., 2010). This aphidicolin and DAPI induced common fragile site was mapped to the terminal region of the short arm of chromosome 7 (7p22) (Yunis and Soreng, 1987). In detail, the DNA sequence of the fragile site was determined using BAC and PAC clones and fluorescent *in situ* hybridization (FISH) analysis. The molecular composition of the identified sequence was also analyzed and searched for the presence of high DNA helix flexibility regions.

To investigate the relationship between DNA replication time and fragility, in the second part of my work I have investigated the timing of replication of sequences mapping within two DAPI inducible fragile sites FRA1H and FRA2G by using FISH on interphase nuclei (Pelliccia et al., 2008).

The last part of my work focused on the role of common fragile sites in the generation of genome instability. There is extensive evidence that common fragile sites are preferential loci for double strand breaks under stressful growing conditions, which can result in gene amplification as well as deletions and translocation. Myllykangas et al. (2006) performed a bibliomic survey to investigate DNA amplifications in different neoplasms. They identified frequently amplified chromosomal loci, and found that the amplifications are preferentially localized in certain genome regions and contain many cancer genes and fragile sites. However, also according to these Authors, the demonstration of the colocalization of amplification breakpoint regions and fragile DNA sequences needs of studies using molecular biology resolution. In the model of intrachromosomal gene amplification obtained by breaks at common fragile sites, the amplicons appear to be arranged in head-to-head organized ladders created by repeated breakage-fusion-bridge cycles (Coquelle et al., 1997, see introduction and Fig. 3). Two common fragile sites may set the boundaries of the early amplicon: the break at the more telomeric site determining the telomeric loss and then the sister chromatids fusion, and the break at the more centromeric site defining the amplicon size. Common fragile sites FRA1H, FRA2G and the latest characterized FRA7B, are frequently breakpoints in chromosome aberrations in various neoplasms. At a distance of about 20 Mb from FRA1H, at the chromosome band 1q44, the common fragile site FRA1I is mapped, and about 15 Mb from FRA2G, at the chromosome band 2q32.1, was mapped FRA2H. FRA7B is mapped at the end of the short arm of chromosome 7, in this case a telomeric breakage could be involved in giving rise

to a BFB cycle. In fact, telomere dysfunction was associated to fusion-breakage driving to amplifications and deletions of cancer-relevant loci, suggesting that telomere-based crisis provides a mechanism of chromosomal instability (O'Hagan et al., 2002).

Therefore, insofar as the entire 1q arm, the chromosome region 2q13-q36 and 7p13-22 had been included in the top 30 identified amplification hotspots (Myllykangas et al., 2006), a panel of 21 cancer cell lines were analyzed to identify any DNA copy number variation affecting the region on chromosome 1 and 2, between adjacent common fragile sites, and the region of chromosome 7 between FRA7B and the telomeric sequences (Pelliccia et al., 2010).

The work was performed using FISH and BAC probes chosen in the region between fragile sites.

RESULTS AND DISCUSSION

PART 1. Molecular characterization of the human common fragile site FRA7B

The FRA7B common fragile site is located at the terminal region of the short arm of chromosome 7 in such a way that the telomeric region is sometimes observed as being completely undercondensed (Fig. 9).

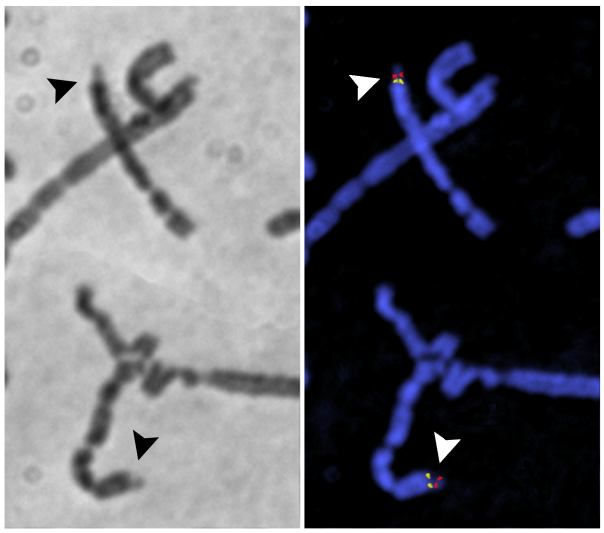


Figure 9. GIEMSA stained metaphases chromosomes expressing the fragile site FRA7B (arrowheads). Hybridization signals of the BAC clones RP11-42B7 (red) and RP11-425P5 (yellow) on the same chromosomes stained with DAPI.

To analyze the fragile region and to define the centromeric boundary of the FRA7B sequence, a set of nine PAC and BAC clones were used for FISH analysis. The probes were selected starting from a sub-telomeric region and walking toward the centromere (Tab. 3)

The undercondensed area sometimes extends as far as the telomere, and the telomeric probe sometimes spreads on the undercondensed region. Also, since telomeres are regions with replication problems that can display an increase in frequency of a fragile site-phenotype after aphidicolin treatment in metaphase (Sfeir et al. 2009), the question could arise as to whether it is correct to consider the $(TTAGGG)_n$ telomere sequence (a few kilobases long) as a part of the FRA7B fragile site.

Table 3. Analysis of FISH signals of BAC and PAC clones and of a telomere probe relative to FRA7B fragile region. In gray is underlined the FRA7B more fragile region.

Probe	N	% Prox	% Mid	% Dist
RP11-708O1	69	100	0.0	0.0
RP11-507C1	71	98.6	1.4	0.0
RP4-755G17	42	90.5	2.0	7.5
RP11-696N1	73	80.8	4.1	15.1
RP11-425P5	85	54.7	5.9	39.4
RP11-42B7	75	22.7	13.3	64.0
RP11-151M24	50	8.0	22.0	70.0
RP11-6A1	52	5.7	23.1	71.2
RP11-449P15	49	4.1	20.4	75.5
(TTAGGG)n	90	0.0	18.0	80.0

%Prox, %Mid and %Dist refer to the percentage of time that the hybridization signal was observed proximal to, on, and distal to the gap/break; N is the number of FISH observations for the probe.

The fluorescent signal of the telomeric $(TTAGGG)_n$ probe was observed as a distal (telomeric) spot to the FRA7B gap/breakage or to spread over the undercondensed region. The hybridization signals of the seven BAC were observed to lie proximal to, on, and distal to the FRA7B gap/breakage on different chromosomes (Fig. 9, Tab. 3). They are therefore located inside the fragile region. None of the fluorescent signal of the BAC clone RP11-507C1 was distal to the FRA7B gap (98% of proximal hybridization signals), while the fluorescent signals of the BAC clone RP11-708O1 were completely proximal (centromeric) to the FRA7B gap/breakage, so this last BAC is not involved in the fragility (Tab. 3). Inside the fragility region, a more fragile region may be identified around and above the BAC clone RP11-425P5. This probe presents hybridization signals both proximal and distal to the FRA7B gap/breakage with high frequency (Tab 3, Fig. 10). The fragile region FRA7B, from the BAC 507C1 to the telomere, is 12.2 Mb long.

The FRA7B sequence was then analyzed to identify coding sequences and some structural features possibly involved in fragility. FRA7B spans the distal part (~5 Mb) of the G-band 7p21.3 and the entire band 7p22, which is subdivided at the 850-band level of resolution, into the R-band 7p22.1, the light gray G-band 7p22.2, and the R-band 7p22.3 (Fig. 10).

Using public databases, the FRA7B DNA sequence was analyzed to identify coding sequences, AT content, DNA repeats, CpG islands, MAR, high flexibility regions. Following the NCBI database (Build 36.3), the FRA7B sequence was divided into four parts: the distal 5 Mb of the G-band 7p21.3 and the three sub-bands of 7p22 (Fig. 10, Tab. 4). The same analysis was performed on four non-fragile control sequences with GC content similar to each of the four fragile sequences, obtained from G-bands (NFRA-G) and R-bands (NFRA-R) (Tab. 4).

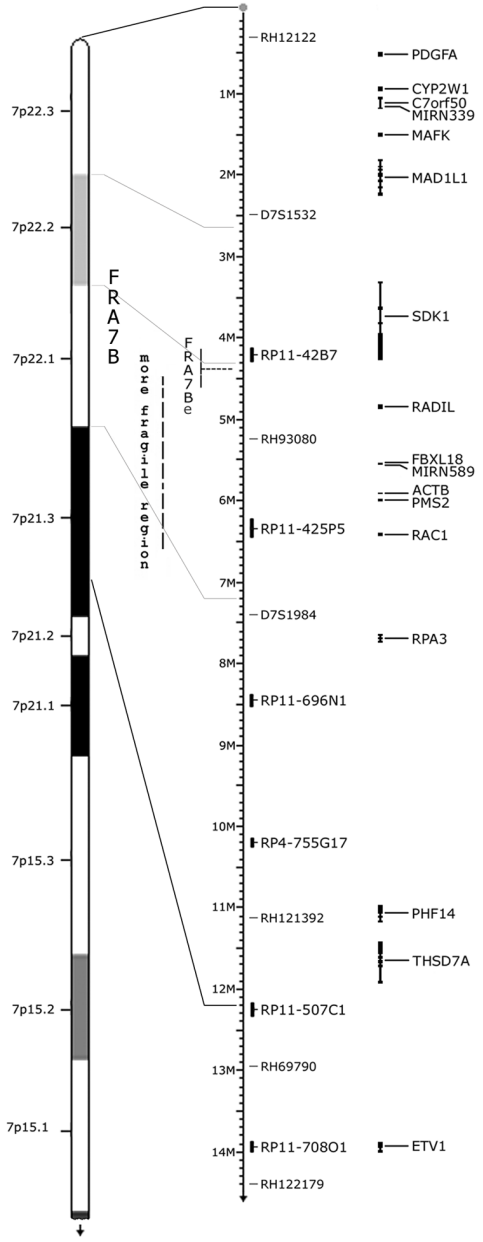


Figure 10. Physical map of the FRA7B region. The BAC and PAC clones used as probes for its characterization, some markers and some genes are shown.

Table 4. Chromosomal position of the analyzed fragile and nonfragile control sequences located in G- and R-bands (Build 36.3).

7p21.3 G-b	7.2-12.2 Mb	5 Mb
NFRA-Ga	8p22 (16.3-17.3 Mb)	1 Mb
	12p12.3 (15-16 Mb)	1 Mb
	13q31.1 (83-84 Mb)	1 Mb
7p22.1 R-b	4.3-7.2 Mb	2.9 Mb
NFRA-Rb	16p13.3 (4.4-5.9 Mb)	1.5 Mb
	12p13.31 (5.0-6.5 Mb)	1.5 Mb
7p22.2 G-b	2.65-4.3 Mb	1.65 Mb
NFRA-Gc	12p13.32 (3.3-4.8 Mb)	1.5 Mb
7p22.3 R-b	0-2.65 Mb	2.65 Mb
NFRA-Rd	9q34.3 (137.2-138.2 Mb)	1 Mb
	16p13.3 (1.4-2.9 Mb)	1.5 Mb
FRA7Be R-b	7p22.2 (4370.22-4460.22 kb)	0.09 Mb
NFRA-Re	12p13.32 (5910-6000 kb)	0.09 Mb

First, the sequences were analyzed, 500 kb at a time, and the average values of AT content, DNA repeats, CpG islands, MAR, high flexibility regions within the four regions of interest and the control region were then computed.

Sequence analysis of characterized common fragile sites has revealed that very large genes are associated with approximately half of the fragile regions (Smith et al., 2006). Moreover, miRNA genes, a family of small non-coding RNAs involved in gene regulation, have frequently been found located at fragile sites (Calin et al., 2004). A total of 74 genes are mapped at the fragile region. Two of them are miRNA genes: *MIRN589* (99 bp), codified within the third intron sequence of the gene *FBXL18* (F-box and leucine-rich repeat protein 18) in the R-band 7p22.1, and *MIRN339* (94 bp), codified within the

second intron of the gene *C7orf50* (Chromosome 7 open reading frame 50) in the R-band 7p22.3. Moreover, three are large genes: *THSD7A* (thrombospondin, type 1, domain containing 7A) (458 kb) mapped at G-band 7p21.3, the very large gene *SDK1* (side-kick homolog 1, cell adhesion molecule) (968 kb) mapped at the G-band 7p22.2, and *MAD1L1* (mitotic arrest deficient-like 1) (418 kb) at R-band 7p22.3 (Fig. 10). All these genes regulate important cellular functions. In particular, the miR-339 is involved in promoting resistance of cancer cells to cytotoxic T-lymphocytes (Ueda et al., 2009). Moreover, *MAD1L1* is a component of the mitotic spindle-assembly checkpoint, and its dysfunction is associated with chromosomal instability and pathogenesis in various type of cancer (Tsukasaki et al., 2001).

The fragile sequence in the G-band 7p21.3 (5 Mb) has an AT average content, 63.6% ± 1.4 ($P \leq 0.05$), that represents a very high value for G-bands. The AT average content of the R-band 7p22.1 (3 Mb) is 51.9% ± 3.9 ($P \leq 0.05$), which is a very high value for R-bands. The light G-band 7p22.2 (1.65 Mb) has a very low AT average content, 54.7% ± 4.4 ($P \leq 0.05$), for G-bands. The average content of the R-band 7p22.3 (2.65 Mb), 45.1% ± 2.3 ($P \leq 0.05$), represents a median value for R-bands (The BAC resource Consortium, 2001; Niimura and Gojobori, 2002).

The search for putative CpG islands in the fragile regions shows quantities of these sequences that are not significantly different from that of the control sequences.

An analysis of the DNA repeat composition of the four FRA7B sequences does not display any important differences between these and the NFRA-G and NFRA-R control sequences (Fig. 11), except for the over-representation of Alu elements in the R-band 7p22.1. This difference is also observed in the comparison with standard genome sequences with similar AT content (Smit, 1999).

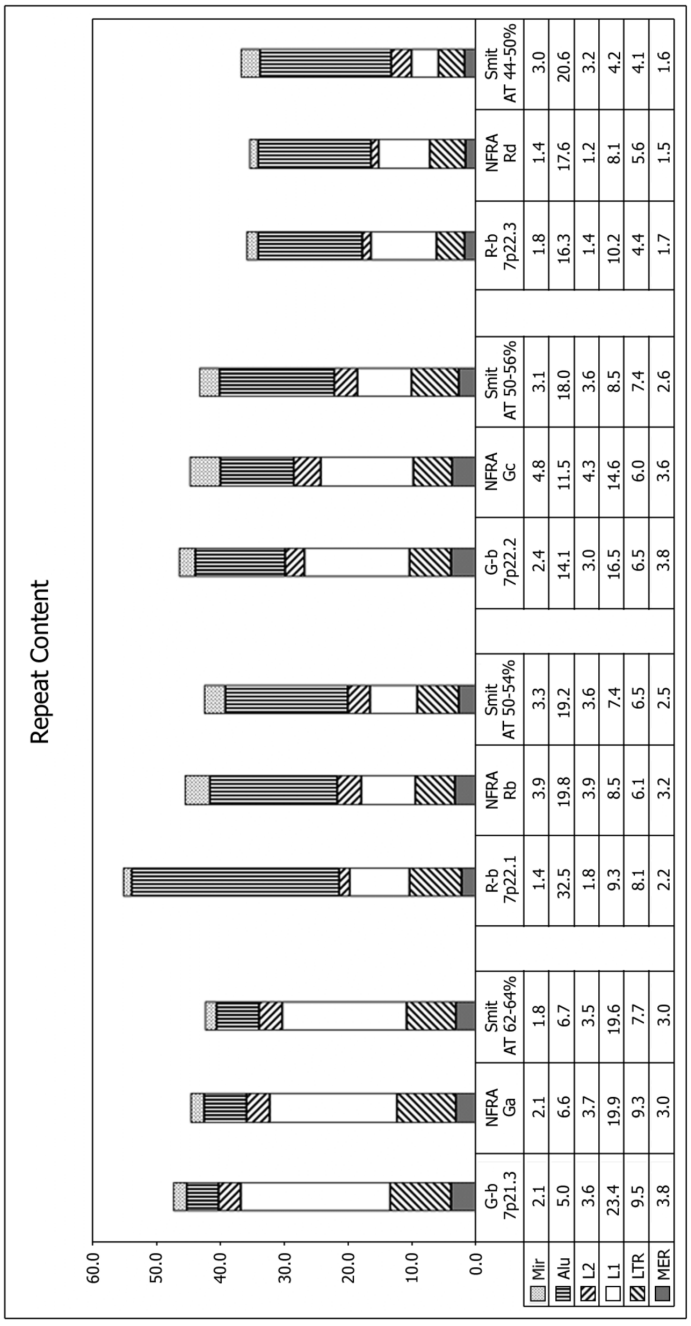


Figure 11. Sequence analysis of fragile and non fragile control regions.

No significant difference in the density of potential MAR sequences is shown in the fragile regions in comparison with the control NFRA regions.

When the four FRA7B regions were analyzed for the presence of DNA helix high-flexibility regions, no significant differences were observed between the fragile and non fragile control sequences, although this analysis pointed to the presence of a particular region, FRA7Be, about 90 kb long (Tab. 5, Fig. 10).

Table 5. Twistflex flexibility analysis of fragile and nonfragile control regions (curvature threshold 13.7)

Bands	Mb	Flexibility (peaks/Mb)	Unified peaks[a]/Mb	Cluster of peaks[b]/Mb
7p21.3 G-b	5.0	79	41	10.4
NFRA-Ga	3.0	71	44	6.7
7p22.1 R-b	2.9	40	24	5.5
NFRA-Rb	3.0	42	21	3.0
7p22.2 G-b	1.65	41	24	2.4
NFRA-Gc	1.5	30	18	3.6
7p22.3 R-b	2.65	17	11	2.4
NFRA-Rd	2.5	16	7	1.2
FRA7Be R-b	0.09	378	133	55.6
NFRA-Re	0.09	156	67	11.1

[a] Group of peaks with distance between adjacent peaks ≤ 100bp
[b] Three or more peaks, with distance between adjacent peaks ≤ 5kb

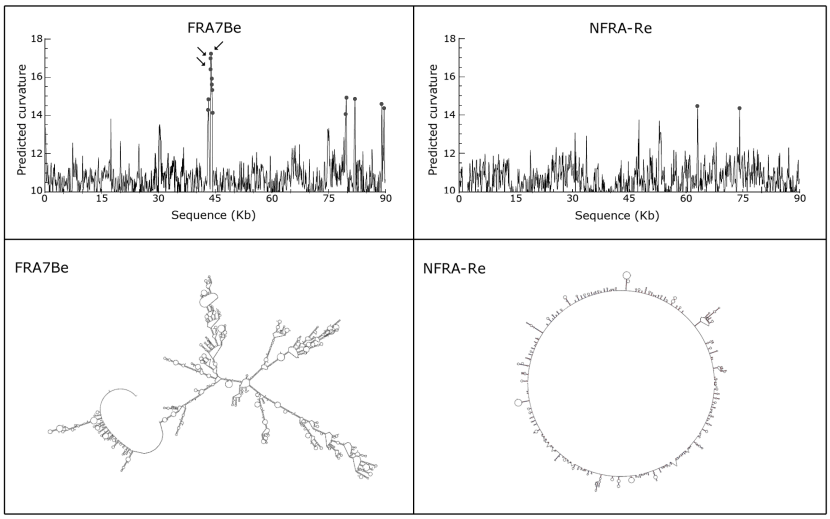
FRA7Be is localized at the telomeric edge of the 7p22.1 R-band, just at the beginning of the more fragile region, and presents very high flexibility values also when compared with a non fragile control sequence (NFRA-Re) chosen as the richest in flexibility peaks from a number of examined control regions having the same length and AT content (Tab. 5, Fig. 12). This sequence, which could be considered AT rich because of its localization in an R-band, is not particularly AT rich (56.8%). However, its AT content is due overall to runs of AT-rich simple repeats that coincide with the flexibility peaks.

Using the M-fold program, the FRA7Be region was analyzed for its ability to form ssDNA secondary structures. FRA7Be is extremely favorable for forming stable secondary structure while the NFRAe control region doesn't show the same propensity (Fig. 12).

In this part of the work, the extension of the common fragile site FRA7B has been defined and some of its molecular features have been analyzed. The cytogenetic location of the FRA7B fragile sequence occupies most of the G-band 7p21.3 and the whole telomeric 7p22 band (subdivided into the three sub-bands described at the 850-band level of resolution: the R-band 7p22.1, the light gray G-band 7p22.2, and the R-band 7p22.3). So, like most common fragile sites (El Achkar et al., 2005), FRA7B lies at the interface of G- and R-bands. These regions are probably difficult to replicate because of the transition from late to early replicating DNA domains. A large body of evidence supports the replication-dependent instability of the common fragile regions. Chan et al. (2009) recently found that replication inhibitors such as aphidicolin induce sister chromatid DNA bridges as replication intermediates, specifically at common fragile site loci that are already regions intrinsically difficult to replicate. These pathologic structures, inefficiently

resolved before anaphase by DNA repair protein complexes, will lead to anaphase bridges, then to chromosome breakages, and so forth, to all events derived from these breakages, such as chromosome rearrangements, recombination, viral integrations, and amplifications.

Figure 12. Analysis of DNA helix flexibility of FRA7Be and NFRAe control sequence. The flexibility peaks with curvature values above 14 are marked by dots. Values of curvature above 16 are marked by arrows. The horizontal axis indicates the size of the analyzed sequences, the vertical axis shows the curvature values in the DNA twist angle.
Below are presented the M-fold analysis of the FRA7Be and NFRAe high flexibility regions.



Sequence analysis of a number of common fragile sites seems to indicate that the replication fork progression may be disturbed by highly flexible DNA sequences, so that the latter are believed to have an important role in fragility (Zlotorynski et al., 2003; Mishmar et al., 1998). Flexibility analysis of the FRA7B region revealed the presence of a 90 kb long sequence, FRA7Be, particularly rich in AT DNA stretches and in flexibility peaks right at the beginning of the more fragile region and localized at the telomeric edge of the 7p22.1 R-band. This sequence could have a role in inducing stalling of the replication and fragility in the FRA7B region due to its propensity to form stable secondary structures.

The 7p22.1 R-band, besides harboring the more fragile region and the FRA7Be flexible sequence, also displays a particular richness in Alu repeats element. It is not known whether there is any link between Alu richness and fragility. Nonetheless, even if it is debatable whether Alu repeat richness is the cause or the consequence of fragility, the R-band 7p22.1 possesses particular features that could have an important role in generating instability in this chromosomal region.

PART 2. Replication timing of two human common fragile sites: FRA1H and FRA2G

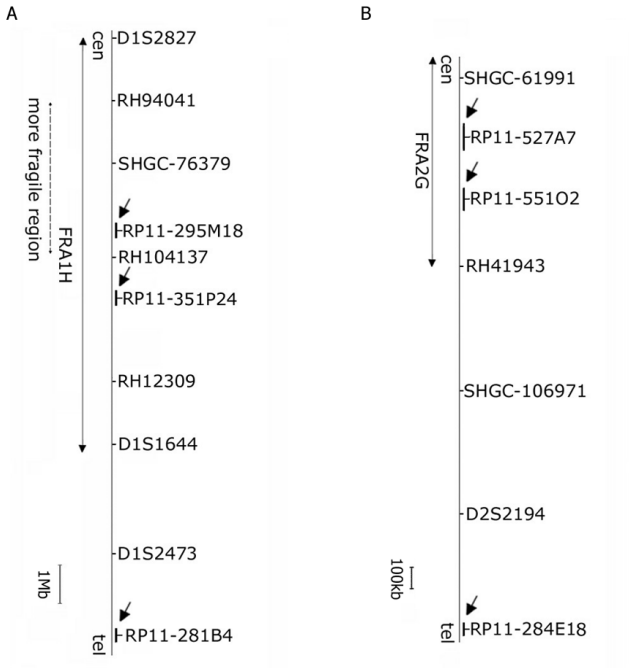
Starting from the assumption that common fragile sites are regions of late or slowed replication in the presence of sequence elements that have the propensity to form secondary structures, and that the cytogenetic expression of fragile sites may be due to unreplicated DNA, this part of my work analyzes the relationship between DNA replication time and fragility.

In particular the timing of replication of sequences mapping within two common fragile sites, FRA1H and FRA2G was investigated. These two fragile sites are both efficiently induced by DAPI. FRA1H (1q41-q42.1) spans about 10 Mb, and its expression is induced also by 5-azaC, 5-azadC, and by ad12 (Curatolo et al., 2007). FRA2G site (2q24.3-q31) extends over 1 Mb, and belongs to the aphidicolin induced fragile sites class (Limongi et al., 2003). Moreover, the timing of replication of syntenic non fragile sequences, and that of early and late replicating control sequences were analyzed by using fluorescent in situ hybridization on interphase nuclei.

In this method, before replication, each locus appears as a single hybridization dot signal (S) while, after replication, duplicated dot signals (D) are visible (Selig et al., 1992).

The probes used to study the replication timing of the FRA1H site were the BACs RP11-295M18 (FRA1Hfa) and RP11-351P24 (FRA1Hfb), which map within the fragile region. Moreover, in order to directly compare the replication time of the fragile sequences with a non fragile sequence localized on the same chromosome, cells were also hybridized with the BAC clone RP11-281B4 (NFRA1), which is located outside the fragile region (Fig. 13A) (Curatolo et al., 2007).

Figure 13. Maps of the FRA1H (A) and FRA2G (B) regions. Some DNA markers are shown. Arrows indicate the BAC used for the replication timing analysis by FISH.



The probes used to study the replication timing of the FRA2G site were the BACs RP11-527A7 (FRA2Gfa) and RP11-551O2 (FRA2Gfb), which map within the fragile region. Moreover, cells were also hybridized with the BAC clone RP11-284E18 (NFRA2), which is located outside the fragile region (Fig. 13B) (Limongi et al., 2003).

The interphase nuclei in S-phase were distinguished from the G1- and G2-nuclei by adding BrdU 20 min prior to cell harvest. Five distinct temporal S-phase stages, from early (stage I) to late (stage V) S phase, were identified analyzing the BrdU-labelling patterns of the nuclear DNA replication (Fig. 14) (O'Keefe et al., 1992; Sadoni et al., 1999).

Two experimental protocols were performed for both FRA1H and FRA2G fragile site replication analysis. In the first type of experiment, the nuclei were observed by an epifluorescence-equipped microscope and all probes were hybridized in pairs and simultaneously with either the GG1 or Me2.5 cosmids used as early and late replicating control probes, respectively (Morley et al., 1991; Smith and Higgs, 1999). The use of these control probes with the probes of interest also allowed checking the comparability of results.

The frequency of duplicated signals (D) of each probe was estimated for each of the five nuclear BrdU-labeling pattern stages. At least 200 S-phase nuclei were scored for each probe. The efficiency of hybridization was controlled by scoring the hybridization of metaphase chromosomes on each analyzed slide.

In the second type of experiment, all probes were individually hybridized and the observations were obtained by confocal microscopy in order to identify both the BrdU nuclear and the probe-labeling pattern by three-dimensional analysis. Also in this case, at least 200 S phase nuclei were scored for each probe and the frequency of duplicated signals (D) was estimated for each nuclear DNA replicating pattern.

The frequencies of D signals observed for each probe and for each of the five temporal replicating nuclear stages, obtained with the two experimental methods, did not present statistically significant differences and therefore the observations were combined.

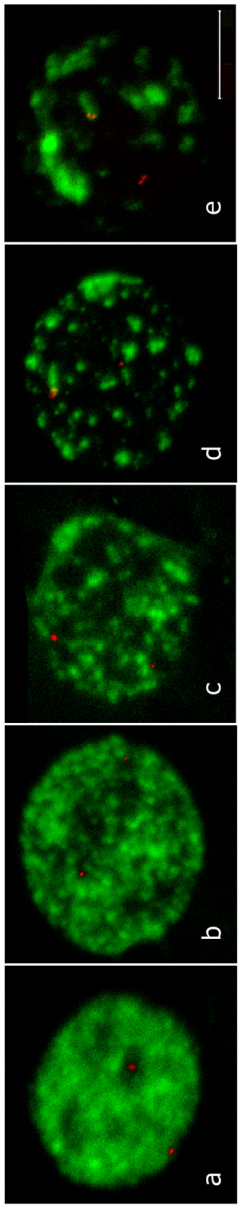


Figure 14. Human lymphocytes BrdU replication labeling patterns (green) of five temporal, from early I (a) to late V (e) S phase stages. The images were obtained from single midnuclear sections by confocal microscopy analysis. The FISH signals (in red red) refer to the FRA1Hfa probe. They were often revealed on different sections and then merged. Bar 10 μm.

All the experiments were performed in duplicate with and without addition of DAPI. The aim was to study the possible effect of DAPI, a compound that induces undercondensation of FRA1H and FRA2G, on replication of these two sites. No statistically significant differences were found between the two treatment types.

The results are presented in Fig. 15. As regards the FRA1H site, the analysis of the frequency of D signals shows that both the FRA1Hf fragile sequences and the NFRA1 non fragile sequence initiate their duplication early at stage I of the S phase. Nevertheless the proportion of doublets at this stage and at the following stages is very different for the fragile and non fragile loci. FRA1Hfa and FRA1Hfb present only a few D signals (about 6% and 11% respectively) at stage I and the number of D signals increases up to about 55% and 50%, respectively at stage V. NFRA1 is already duplicated at stage I in about 21% of the cells, then the frequency of D signals increases up to 100% at stage V (Fig. 15).

Likewise, as regards the FRA2G site, both the replication of the fragile FRA2Gf sequences and that of the non fragile NFRA2 sequence initiates early at stage I, but while FRA2Gfa and FRA2Gfb present only about 10% and 11%, respectively of D signals, NFRA2 presents about 32%. At stage V of the S phase, the proportion of doublets reaches about 43% and 57% in the two fragile loci, respectively and 95% in the non fragile locus (Fig. 15).

As expected, GG1 (the α-globin locus) replicates early in S phase (about 26% of D signals at stage I) and is almost completely replicated at stage V presenting about 87% of D signals. Also the replication of the late replicating probe Me2.5 (β-globin locus) initiates at stage I of the S phase, but at this stage shows only about 2% of doublets. Moreover, at stage V, the proportion of D signals reaches only about 50% in this locus (Fig. 15).

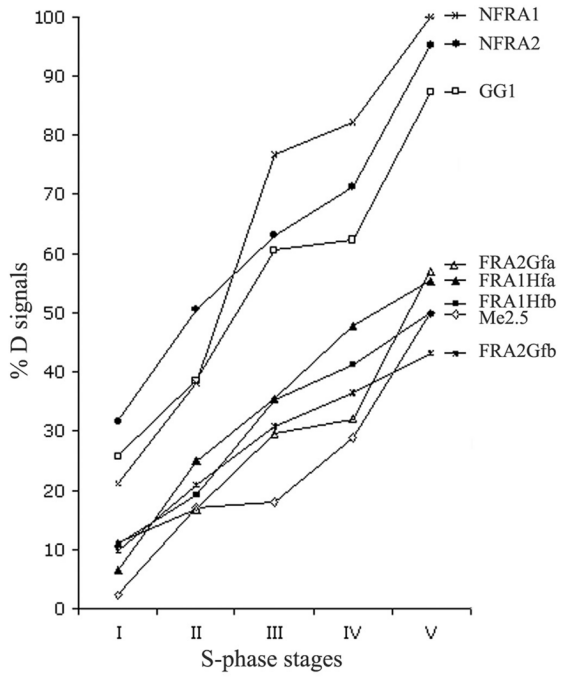


Figure 15. Percentage of double FISH dots (D) observed at the S phase temporal stages (I-V) for the eight probes used.

The comparative analysis of this data reveals that the replication of all examined loci initiates already in the earliest stage of S phase (stage I) but with a very different frequency in each analyzed locus. In fact, the four fragile sequence FRA1Hfa, FRA1Hfb, FRA2Gfa and FRA2Gfb, and the late replicating control sequence Me2.5 show a frequency of doublets ≤ 11% (about 6%, 11%, 10%, 11% and 2% respectively), while the

frequency of D signals of the two non fragile sequences, NFRA1 and NFRA2, and of the early replicating control sequence GG1 is ≥ 21% (about 21%, 32% and 26% respectively). Moreover, the first five loci, FRA1Hfa and b, FRA2Gfa and b, and Me2.5, reach a maximum frequency of duplicated loci of 57% at stage V (about 55%, 50%, 57%, 43% and 50% respectively), while at this same stage NFRA1, NFRA2 and GG1 are completely or in large parts duplicated (about 100%, 95% and 87% respectively). Hence, the fragile sequences and the late replicating sequence enter G2 phase unreplicated in about 50% of alleles.

Laird et al. (1987) first proposed that fragile sites replicate late in the cell cycle 'perhaps extending into what is conventionally called G2', failing the normal condensation and are therefore present as chromosome gaps in metaphase. Moreover, nearly all authors studying the replication timing of fragile sites concluded that it might be completed late in the cell cycle. It has been demonstrated that the replication of a small percentage of the genome occurs during the G2 phase (Widrow et al., 1998). It has also been demonstrated that checkpoint mechanisms exist which ensure that mitosis does not initiate before the replication of the genome is completed. However, low levels of unreplicated DNA may not be sufficient to hinder some cells entering mitosis (Widrow et al., 1998). Therefore very late replicating regions could sometimes reach mitosis unreplicated and be expressed as gaps/breaks. Alternatively, because of G2-replication, these regions could be so late in condensing as to sometimes appear as chromosome gaps/breaks.

However, because not all late replicating regions are fragile sites, at least another structural and/or functional feature besides a late or slowed replication must be shared among fragile sequences. As mentioned above, many researchers assume that the presence of sequence elements potentially able to form secondary structures and to induce stalling of replication forks is a second factor promoting fragility (Arlt et al., 2003; Glover et al., 2005 and references cited therein). On the other

hand, high numbers of flexible sequences rich in AT-dinucleotides able to form secondary structures have been found in most cloned common fragile sites including FRA1H and FRA2G (Limongi et al., 2003; Curatolo et al., 2007).

The inducers of fragile sites may amplify the effects of late replication. Aphidicolin may induce the expression of bulk of common fragile sites by inhibiting the DNA polymerases. In this case, both FRA1H and FRA2G are induced by DAPI as well as other compounds (see introduction). DAPI preferentially binds to AT-rich DNA and acts in G2 phase as an undercondensing agent (Prantera et al., 1981). The combination of late replication with the presence of sequences receptive to DAPI may make some fragile sites available for binding with this compound and thus no longer capable of undergoing normal condensation.

PART 3. Fragility and genome instability

The last part of the work focused on the consequence of fragile site instability in the genome, in particular in stressful growing cell conditions, such as cancer cells.

Common fragile sites are preferential loci for double strand breaks, which make them important factors in generating genome instability. This instability can result in gene amplification, as well as chromosome deletions and viral insertions (see introduction).

The amplification of genome regions is often observed in human tumors probably because it contributes to the overdosage of oncogenes, implicated by the pathology present in the amplified region.

Many researchers have indicated common fragile sites as sites of chromosome breakages at the origin of breakage-fusion-

bridge (BFB) cycles. As mentioned above, the BFB cycle can be initiated by breakage in one fragile site leading to the telomeric loss, and then the deletion, of a chromosome region. The second breakage that occurs after chromatid fusion, can also happen within a second fragile site that sets the amplicon size.

In this part of the work, the genome of 21 cancer cell lines (see Material and Methods) were analyzed for the presence of DNA copy number variation with breakpoints in the fragile sites FRA1H and FRA1I on chromosome 1, the fragile sites FRA2G and FRA2H on chromosome 2, and the fragile site FRA7B and the telomere on chromosome 7.

According to Myllykangas et al. (2006) the entire 1q arm, the chromosome region 2q12-q36, and the 7p22 region, have been identified as amplification hot spots in several cancers.

Using FISH analysis to investigate the presence of amplification/deletions affecting the chromosome region between the common fragile sites FRA1H and FRA1I and the one centromeric to FRA1H, a BAC clone (RP11-385F5) located between the two fragile sites, and a BAC clone (RP11-434B7) located centromeric to FRA1H were used as FISH probes. FISH of the two probes on metaphase spreads of 21 tumor-derived cell lines (see Material and Methods) did not reveal any amplification nor deletions.

To investigate the presence of amplification/deletions affecting the chromosome region between the common fragile site FRA7B and the telomere and the one centromeric to FRA7B, a BAC clone (RP11-42B7) located between the fragile site and the chromosome end, and a BAC clone (RP11-708O1) located centromeric to FRA7B were used as FISH probes. FISH of the two probes on metaphase spreads of 21 tumor-derived cell lines (see Material and Methods) did not revealed any amplification nor deletions.

To investigate the regions between FRA2G and FRA2H and the one centromeric to FRA2G, the BAC clones RP11-

158L8 located between the two fragile sites and the RP11-681B22 located centromeric to FRA2G were used (Fig. 16). FISH with these two probes on the panel of 21 cancer cell lines revealed the existence of a duplicated region on a chromosome der(2) in two analyzed leukemia cell lines K562 (K562-1 and K562-2). The duplications on the two chromosomes der(2) of the two cell lines are of different lengths.

The K562 cell line derives from a pleural effusion of a female with chronic myelogenous leukemia (CML) in terminal blast crisis (Lozzio and Lozzio, 1975). Different sublines that differ slightly in their marker set probably exist. The hallmark of CML is the Philadelphia rearrangement of the *BCR* and *ABL* genes [t(9;22)(q34;q11)] which is responsible for the *BCR/ABL* fusion gene and the expressed fusion protein that is critical in the neoplastic transformation (Naumann et al., 2001).

In the **K562-1** cell line, the probes 158L8 and 681B22 showed a normal hybridization pattern on two chromosomes 2, while on the long arm of a third chromosome [der(2)-1] duplicated signals appeared in the following order: centromere-681B22-158L8-158L8-681B22-telomere (Figs. 17A and 18A).

To define the size of the duplicated region and to precisely identify its boundaries and thus the breakpoints involved in its origin, FISH experiments were performed using two sets of BACs, one set distal (telomeric) to 158L8 (335G13, 489J10, 363G3) and the other set proximal (centromeric) to 681B22 (472M4, 341F20, 375H16, 364H22) (Fig. 16).

With regard to the BAC set distal to 158L8, the analysis of the FISH signals revealed very close fluorescent duplicated dots of the probe 335G13 and the absence of signals for the probes 489J10 and 363G3. These results place the breakpoint and thus the telomeric boundary of the duplication in the region between the sequences covered by the probes 335G13 and 489J10 in the band 2q32.1 (Fig 18A). The common fragile site FRA2H is mapped in this band (Fig. 16).

Figure 16. Map of part of the long arm of chromosome 2 showing the BAC cloned used as probes for FISH experiment, and the localization of FRA2H, FRA2G and FRA2S.

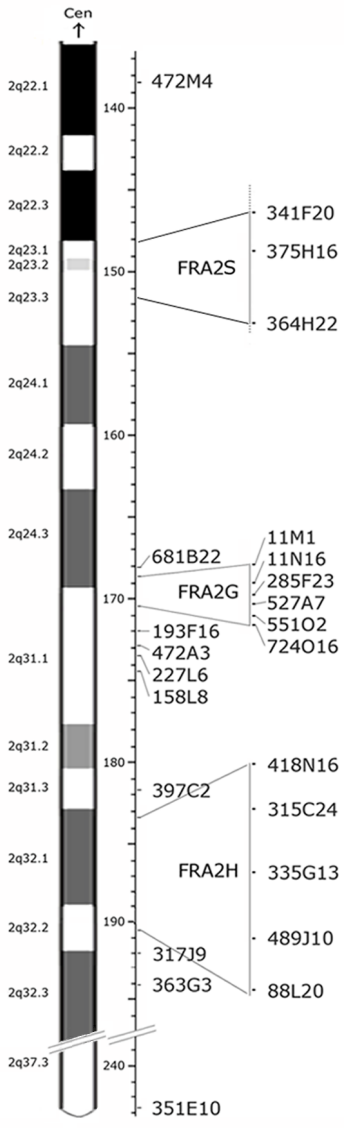
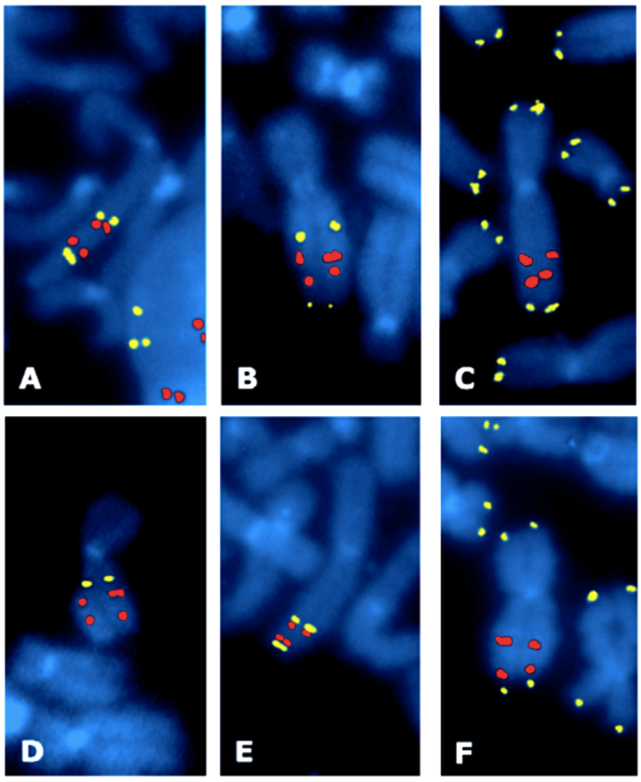
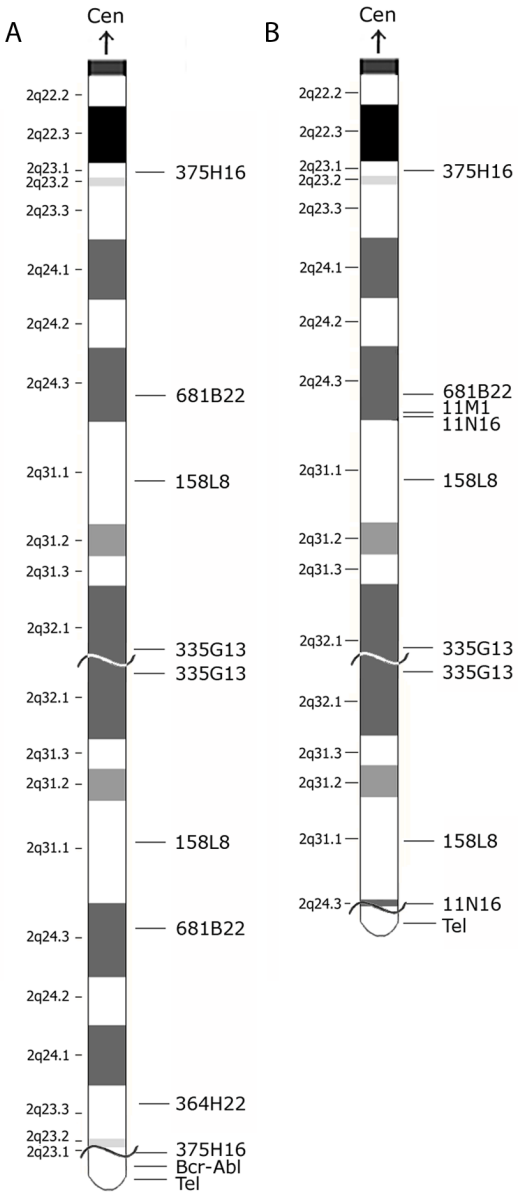


Figure 17. Some of the FISH experiment showing the existence of the inverted duplication on the chromosome der(2)-1 (A-C) and der(2)-2 (D-F). (A) the palindromic arrangement of the clones 681B22 (yellow) and 158L8 (red). (B) the palindromic arrangement of the clones 375H16 (yellow) and 158L8 (red), note moreover that the telomeric fluorescent signal of the probe 375H16 is consistently less intense than the centromeric signal, indicating that the sequence is partially deleted. (C) the presence of the telomeric signal of the $TTAGGG_n$ probe (yellow) on telomere of the rearranged der(2)-1 identified by the duplicated signals of the 158L8 probe (red).



(D) on the chromosome der(2)-2 only the probe 158L8 (red) displays duplicated signals, while the 681B22 (yellow) has a single signal. (E) the palindromic rearrangement of the clones 11N16 (yellow) and 158L8 (red). (F) the presence of telomeric sequences (yellow) on the der(2)-2 identified with the 11N16 probe (red).

Figure 18. The two chromosome der(2)-1 (A) and der(2)-2 (B) are represented. The two wavy lines indicate the breakpoints at the origin of the two rearranged chromosomes.



With regard to the proximal BAC set (centromeric) to 681B22, fluorescent signals of the probes 472M4, 341F20 in the order and position expected for a normal chromosome 2 were observed, while the probes 364H22 and 375H16 showed fluorescent duplicated signals in the following order: centromere-472M4-341F20-375H16-364H22-364H22-375H16-telomere. Moreover, the telomeric fluorescent signal of probe 375H16 was consistently less intense than the centromeric signal, indicating that is partially deleted (Figs. 17B and 18A). These results place the centromeric boundary of the duplication in the region covered by the clone 375H16 in the band 2q23.1. No common fragile sites have been mapped to this chromosomal region so far.

To investigate the structure of the telomeric region of the chromosome der(2)-1 a telomeric probe $(TTAGGG)_n$ was hybridized to detect the presence of telomeric sequences to cap the rearranged chromosome. This probe produced normal fluorescent signals on the telomeres of all chromosomes, including the der(2)-1 chromosome (Fig. 17C). Hypothesizing that the subtelomeric and the terminal part of the der(2)-1 could arise from reacquiring the sequences originating on the normal chromosome 2, a BAC probe (RP11-351E10) mapping at the subtelomeric region of 2q was hybridized. The signals of this probe were found on the two normal chromosomes 2, but not in the der(2)-1.

Interestingly, using the probes 1132H12 and 835J22, two sequences localized in the *ABL* gene that maps on chromosome 9, and the probe 434O9, a sequence localized in the *BCR* gene that is mapped at the chromosome 22, all these three probes hybridize on the terminal region of the chromosome der(2)-1 (Fig. 19A). This chromosome almost certainly corresponds to the marker chromosome named M1 by Naumann et al. (2001) and possesses the *BCR/ABL* fusion gene.

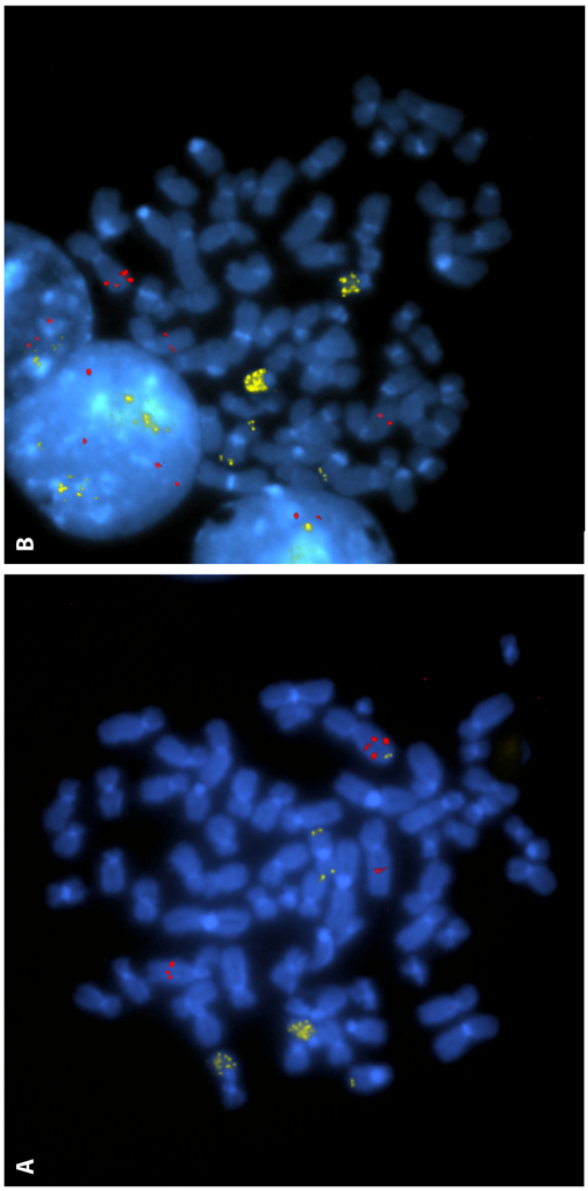


Figure 19. FISH of 1132H12 *ABL* gene probe (yellow) and 158L8 (red) on metaphases from K562-1 (A) and K562-2 (B) cell lines. Note the ABL gene probe marks the terminal region of the chromosome der(2)-1 identified by the double signals of the 158L8 probe.

The three probes 1132H12, 835J22 and 434O9 hybridize with two acrocentric marker chromosomes, which were thus entirely composed of repeated *BCR/ABL* fusion genes, and which certainly correspond to the marker chromosome M20 described by Naumann et al. (2001).

Moreover, as these Authors described, the probes of the *ABL* gene marked also both the telomeric regions of a submetacentric marker chromosome (M7) and one telomeric region of an acrocentric marker chromosome (M8) (Fig. 19A).

It may be concluded from these observation that the der(2)-1 chromosome has an invdup(2)(q23.1q32.1) and a del(2q32.1→tel) (Fig. 18A).

The duplicated region on the der(2)-1 spans about 38 Mb. Sixty-six genes here localized are indicated as possibly implicated in cancer, two genes, *ATF2* (activating transcription factor 2) and *MIRN10B* (microRNA-10b) have been implicated in cancer (Dessen et al., 2010). *MIRN10B*, the expression of which is upregulated in a number of neoplasms, including AML (acute myeloid leukemia), maps to the HOXD cluster, a family of transcription factor genes that play crucial roles during normal development and in oncogenesis.

In the **K562-2** line, the probes 158L8 and 681B22 showed a normal hybridization pattern on two chromosome 2, while on the long arm of a third chromosome [der(2)-2] only the probe 158L8 displayed duplicated signals in the following order: centromere-681B22-158L8-158L8-telomere (Fig. 17D). The palindromic arrangement of the BAC 158L8 suggests that, also in this case, one BFB cycle has occurred and that the break and the subsequent chromatid fusion event are located at a site telomeric to the 158L8 probe, although the absence of a duplicated signal of the probe 681B22 indicates the presence on this chromosome of a narrower duplicated region than that observed in the der(2)-1 of the K562-1 line (Fig. 18B)

To precisely define the duplicated region and to identify the first breakage site, telomeric to the 158L8 probe,

experiments were performed as described for the K562-1 line, yielding the same results. Therefore, also in the der(2)-2 chromosome, the telomeric boundary of the duplication is localized in the region between the sequences covered by the probes 335G13 and 489J10 in the band 2q32.1 in which the common fragile site FRA2H is mapped (Fig.16). To identify the centromeric boundary of the duplication, a new set of probes was used, localized in the region between 681B22 and 158L8 (11M1, 11N16, 527A7, 551O2, 724O16, 193F16, 227L6) (FIg.16).
All the probes, except the probe 11M1, which displayed a single hybridization signal, showed duplicated signals in the following order: centromere-11M1-11N16-527A7-551O2-724O16-193F16-227L6-227L6-193F16-724O16-551O2-527A7-11N16-telomere (Figs. 17E and 18B).

These results place the centromeric boundary of the duplication in the region between the sequences covered by the probes 11M1 and 11N16 in the 2q24.3 band.

From these observations it can be concluded that the der(2)-2 chromosome has an invdup(2)(q24.3q32.1) and a del(2q32.1→tel) (Fig. 18B).

Unlike the der(2)-1 chromosome, no FISH signals were obtained on the terminal region of the chromosome der(2)-2 with the probes 1132H12 and 835J22 (sequences of the *ABL* gene) and 434O9 (sequence of the *BCR* gene). These probes hybridized on other marker chromosomes as described for the K562-1 line (Fig. 19B). The telomeric probe $(TTAGGG)_n$ produced normal fluorescent signals on the telomeres of all chromosomes, including the der(2)-2 chromosome, thus demonstrating that a recapping of the chromosome end has occurred (Fig. 17F).

The duplicated region of the chromosome der(2)-2 spans about 19 Mb and represents a part of the duplicated region on the chromosome der(2)-1.

The chromosome der(2)-2 could be derived from the chromosome der(2)-1 by deletion of the duplicated region owing to a break occurring at 2q24.3 (Fig. 18).

Thirty-eight of the genes localized here are indicated as potentially implicated in cancer. The region contains the two genes, *ATF2* and *MIRN10B*, which have already been implicated in cancer and the HOXD cluster as mentioned above (Dessen et al., 2010).

Except for the chromosome der(2)-2, the analysis of the trypsin G-banded metaphases displays a karyotype very similar to that of K562-1 cell line.

The results assigned to a precise region of the band 2q32.1 the telomeric breakpoint causing the chromatid fusion event of the BFB cycle at the origin of the inverted duplication observed on both chromosomes der(2)-1 and der(2)-2, which are present in the cell lines K562-1 and K562-2 respectively. FRA2H has been mapped at the same band 2q32.1 (Yunis and Soreng, 1984), an aphidicolin-inducible common fragile site yet to be molecularly characterized. To verify that the above-mentioned breakpoint was really located in FRA2H, the sequence of this fragile site was determined using the same approach used to define the fragile sequence of the common fragile site FRA7B (see Part 1).

FISH analysis was performed on metaphase spreads expressing FRA2H with a panel of seven probes (397C2, 418N16, 315C24, 335G13, 489J10, 88L20, 317J9) spanning about 10.3 Mb (Fig. 16), a region somewhat wider than the band 2q32.1. The BAC clone 397C2 gave only proximal hybridization signals while the BAC 317J9 gave only distal signals to the gap/breakage (Tab. 6). Five BAC clones, 418N16, 315C24, 335G13, 489J10, 88L20, showed hybridization signals proximal to, on, and distal to the FRA2H gap/breakage (Tab. 6, Fig. 20); they therefore span this fragile site, which may be estimated as about 8 Mb long.

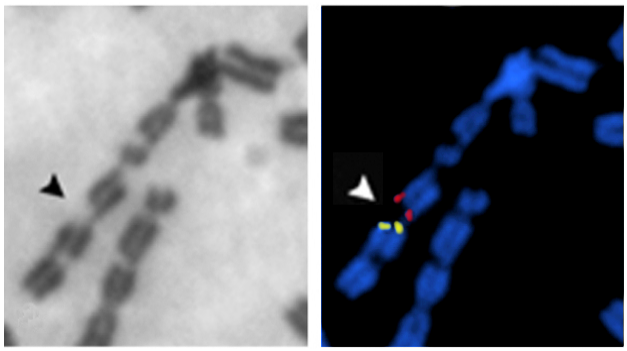
FRA2H spreads over most of the G-band 2q32.1 (6.5 Mb) and over 1.5 Mb of the R-band 2q32.2. Within the fragile region, a more fragile region may be identified between the BAC 315C24 and 335G13, which present hybridization signals both proximal and distal to the FRA2H gap/breakage with high frequency (Tab. 6, Fig. 21).

Table 6. Analysis of FISH signals of BAC clones relative to FRA2H fragile region. In gray is underlined the FRA2H more fragile region.

Probe	N	% Prox	% Mid	% Dist
RP11-397C2	45	100	0.0	0.0
RP11-418N16	50	98.0	2.0	0.0
RP11-315C24	90	63.3	12.2	24.5
RP11-335G13	48	43.8	25.0	31.2
RP11-489J10	58	10.3	6.9	82.8
RP11-88L20	51	5.1	0.0	94.9
RP11-317J9	57	0.0	0.0	100.0

%Prox, %Mid and %Dist refer to the percentage of time that the hybridization signal was observed proximal to, on, and distal to the gap/break; N is the number of FISH observations for the probe.

Figure 20. FISH of the clones 315C24 (red) and 335G13 (yellow) on chromosome 2 expressing FRA2H, the arrowhead indicate the FRA2H break.



These results confirm that the breakpoint, and therefore the telomeric boundary of the duplicated regions of the marker chromosomes der(2)-1 and der(2)-2, which were identified in the region between the sequences spanned by the probes 335G13 and 489J10, is actually localized in the common fragile site FRA2H.

The FRA2H DNA sequence was analyzed in order to identify coding sequences and a number of molecular and structural features (AT content, DNA repeats, CpG islands, MARs, high flexibility regions) possibly involved in fragility.

The fragile sequence was divided into two parts for the analysis, the 6.5 Mb of the G-band 2q32.1 and the 1.5 Mb of the R-band 2q32.2. The same sequence analysis was performed on two non fragile control sequences with GC content similar to each of the two fragile sequences, obtained from G-bands (NFRA-G) and R-bands (NFRA-R) (Tab. 7). The sequences were analyzed 500 kb at a time, and the average values of the two regions of interest and of the control regions were then calculated.

Table 7. Chromosomal position of the analyzed fragile and nonfragile control sequences located in G- and R-bands (Build 36.3).

G-bands		
FRA2H-G	2q32.1 (182.9-189.4 Mb)	6.5 Mb
NFRA-G	8p22 (16.3-17.3 Mb)	1 Mb
	12p12.3 (15.1-16.1 Mb)	1 Mb
	13q31.1 (84.1-85.1 Mb)	1 Mb
FRA2Hf	2q32.1 (185265-185375 kb)	0.11 Mb
NFRAf-G	13q31.3 (83622-83732 kb)	0.11 Mb
R-bands		
FRA2H-R	2q32.2 (189.4-190.9 Mb)	1.5 Mb
NFRA-R	12p12.2 (20.2-21.2 Mb)	1.0 Mb

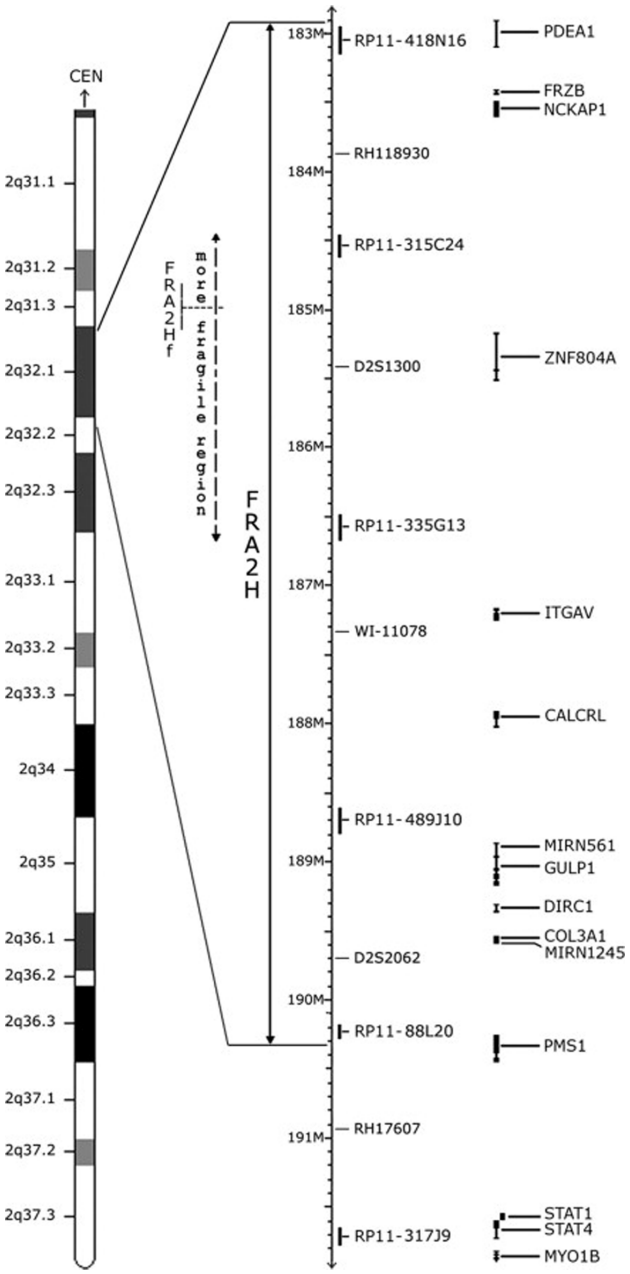


Figure 21. Physical map of the FRA2H region. The BAC clones used as probes for its characterization, some markers and some genes are shown

Twenty-two validated and reviewed genes are mapped in the fragile region. Nine of them are classified as possibly implicated in cancer, while two genes, *DIRC1* (disrupted in renal carcinoma 1) and *PMS1* (postmeiotic segregation increased 1) are indicated as genes implicated in cancer (Dessen et al., 2010). Moreover, two of these genes are miRNA genes: *MIRN589* (97 bp) codified within the first intron of the gene *GULP1* (engulfment adaptor PTB domain containing 1), and *MIRN1245* (70 bp) codified within the first intronic sequence of the gene *COL3A1* (collagen, type 3, alpha 1) (Fig. 21).

The AT average content of the G-band 2q32.1, 64.2 ± 1.3 ($P \leq 0.05$), is a high value for a G-band; and the AT average content of the R-band 2q32.2, 63.8 ± 0.9 ($P \leq 0.05$), represents an unusually high value for an R-band (Niimura and Gojobori, 2002). Investigation of the fragile sequences for putative CpG islands, density of potential MAR, DNA repeat composition, and DNA helix high flexibility regions produced values that were not significantly different from that of the non fragile control sequences.

Nevertheless, this last analysis revealed the presence of a sequence (FRA2Hf) of about 110 kb long, that was very AT-rich (66.6%) and localized in the more fragile region, which presents very high flexibility values (Fig. 22).

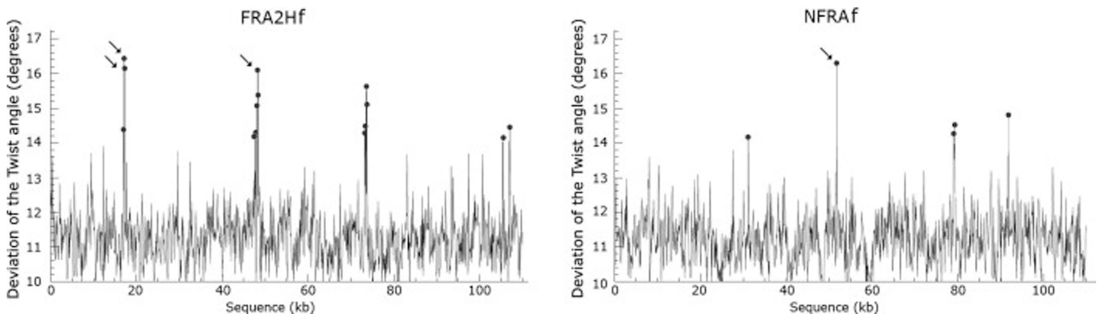
Figure 22. (A) AT content and TwiatFlex analysis of fragile and non fragile control regions (curvature threshold 13.7). (B) Flexibility peaks at FRA2Hf and NFRAf control sequence.

A

Bands	Mb	% AT-bases	Flexibility (peaks/Mb)	Unified (peaks[a]/Mb)	Cluster (peaks[b]/Mb)
2q32.1 G-b	6.5	64.2 ± 1.3	88	51	11.7
NFRA-G	3.0	63.5 ± 1.4	71	44	6.7
2q32.2 Rb	1.5	63.8 ± 0.9	59	39	6.0
NFRA-R	1.0	63.6 ± 0.3	93	45	13.0
FRA2Hf G-b	0.11	66.6	309	173	45.5
NFRAf-G	0.11	66.6	164	109	27.3

[a] Group of peaks with distance between adjacent peaks ⩽100 bp.
[b] Number of peaks ⩾3 with distance between adjacent peaks ⩽5 kb.

B



The results presented above place the centromeric boundary of the invdup(2)(q23.1q32.1), identified in the der(2)-1 chromosome (K562-1 line), in the band 2q23.1, precisely in the region spanned by the clone 375H16. To date, no fragile site has been mapped at this locus. Human lymphocytes from cultures of peripheral blood of five healthy individuals induced by aphidicolin to express fragility, were analyzed by FISH using as probes, besides the BAC clone 375H16, the two BACs 341F20 and 364H22, proximal and distal to 375H16, respectively (Fig.16). The analysis of 150 chromosome gap/breaks (mean = 4.1%; range 2.1-7.2) revealed that in all five individuals analyzed the three probes reside within a fragile region. Indeed, their hybridization signals appeared proximal to, on and distal to a gap/breakage (Fig. 23, Tab. 8).

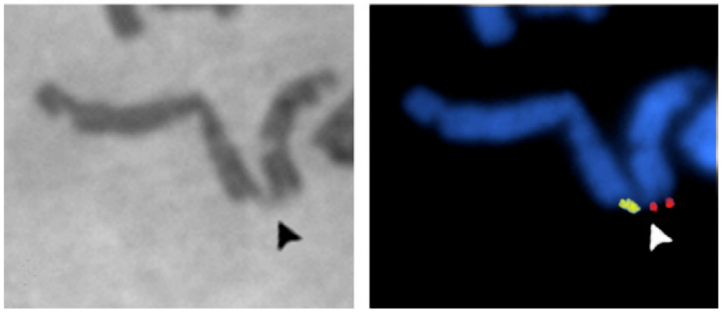


Figure 23. FISH of the clones 364H22 (red) and 341F20 (yellow) on chromosome 2 expressing FRA2S, the arrowhead indicate the FRA2S localization.

Table 8. FISH analysis of BAC clones on chromosomes expressing FRA2S.

Probe	N	% Prox	% Mid	% Dist
RP11-341F20	82	21.9	34.2	43.9
RP11-375H16	69	31.1	31.0	37.9
RP11-364H22	74	8.1	0.0	91.9

%Prox, %Mid and %Dist refer to the percentage of time that the hybridization signal was observed proximal to, on, and distal to the gap/break; N is the number of FISH observations for the probe.

The region delimited by the three BACs used spans about 3.8 Mb. The percentage of gaps/breaks occurring distal to the fluorescent signals of the three probes indicates that the telomeric border of the BAC 364H22 almost defines the distal limit of the fragile region, while the proximal (centromeric) limit is localized centromeric to the BAC 341F20 in the G-band 2q22.3 (Tab. 8, Fig. 23).

This finding indicates that the fragile region identified at 2q22.3-q23.3 is a new aphidicolin inducible common fragile site. Mrasek and colleagues (2010) concomitantly named FRA2S, a breakage localized in the 2q23 region in a screening for aphidicolin-inducible fragile sites. We decided to adopt this alphabetical nomenclature for the common fragile site molecularly mapped by us at 2q22.3-q23.3. At the 2q22.3 band Mattei et al. (1987) mapped the rare fragile site FRA2K by cytogenetic detection.

The centromeric boundary of the invdup(2)(q24.3q32.1), identified in the der(2)-2 chromosome (K562-2 line), in the terminal part of the band 2q24.3, precisely in the region localized between the sequences covered from the BAC clones 11M1 and 11N16, a region that is 100 kb proximal to the common fragile site FRA2G (Limongi et al., 2003). FISH experiments, performed using the probes 11M1 and 11N16 on

chromosomes from cultures of peripheral blood of healthy individuals induced by aphidicolin to express FRA2G (100 gaps/breakages observed), revealed that the fluorescent signals of both the probes were never distal to the gap/breakage but, in small percentages (2% and 4% respectively), on the gap/breakage. This region seems to represent the real proximal part of the FRA2G fragile site, a part that is rarely involved in gap/break events. Thus, this breakpoint can also be localized in a common fragile site.

At a molecular level resolution, these findings add weight to the already substantial evidence that the common fragile sites, as preferential loci for chromosome breakages and telomere loss, have an important role in generating chromosome amplification associated with the neoplasia.

The duplicated regions on the two chromosomes der(2)-1 and der(2)-2, share the telomeric breakpoint and thus the fusion point of sister chromatids, but have two different centromeric boundaries, that leads to different lengths in the amplified region. The three breakpoints are localized in the sequence of three common fragile sites: FRA2H, FRA2S and FRA2G.

CONCLUSIONS

In this work various aspects of common fragile site were analyzed, with particular attention to the family of common fragile sites induced by DAPI.

First, the FRA7B fragile site was molecularly characterized (Bosco et al., 2010). The sequence analysis showed that in this region are mapped many genes that could have a role in neoplastic transformation. Two microRNA genes (miR-589 and miR-339) and *MAD1L1* have already been associated with some types of cancer. It will be interesting to study the pattern of expression of these and other genes of the FRA7B region in different cancer cell lines in order to investigate for homozygous deletions or for modification or loss of gene expression.

The replication analysis of the FRA1H and FRA2G sites showed that these two common fragile site are late/slow replicating (Pelliccia et al., 2008). This replication pattern confirms the hypothesis for the induction of these sites by DAPI. The combination of late replication (due to formation of secondary structures at high flexibility islands) with the presence of sequences receptive to DAPI make this class of fragile sites available for binding with this compound and thus no longer capable of undergoing normal condensation.

In fact, the addition of DAPI does not change the time of replication of these regions since its action is mediated in the G2 phase, during chromosome condensation.

The replication time analyses of the fragile site FRA7B will be done to further confirm this model.

The last part of the work adds weight to the already substantial evidence that the common fragile sites, as preferential loci for chromosome breakages and telomere loss, have an important role in generating chromosome amplifications associated with neoplasia. Here, we have molecularly defined the localization of the breakpoints at the origin of the duplicated chromosome region found on the chromosomes der(2)-1 and der(2)-2 present in two CML cell lines K562-1 and K562-2. The three breakpoints are localized in the sequences of three common fragile sites (Pelliccia et al., 2010).

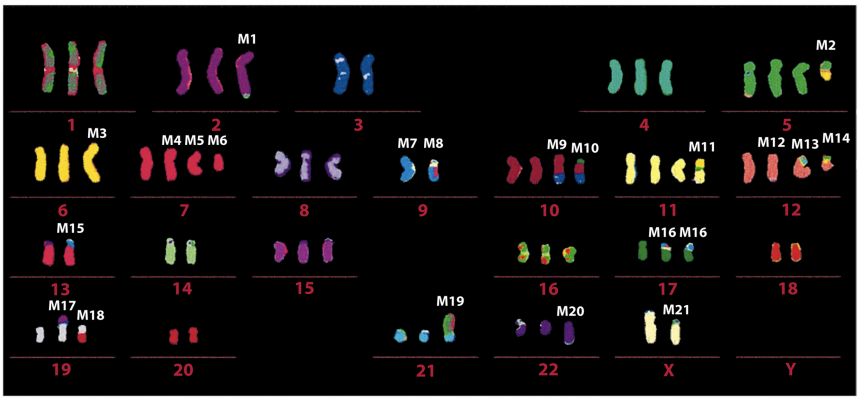
Finally, this work presents FISH as an important multifunctional tool for the analysis of various biological processes, as an indispensable cytogenetic tool to identify specific sequences on chromosomes, and also for other different application, such as the IF-FISH used in this work to analyze the replication timing of genome sequences.

Many modern and sophisticated technique during the past few years have been developed to study chromosome structure and to karyotype cells, such as microarray and comparative genomic hybridization (CGH). But like all the macroscopic approaches to biological processes they lacks of sensitivity. FISH analysis remains a useful tool to study in depth and to precisely identify the breakpoints involved in many different chromosome rearrangements.

For example, the K562 cell line has already been analyzed by Naumann et al. (2001) in a combined analysis of G-banding, multiplex-fluorescent in situ hybridization (M-FISH) and comparative genomic hybridization (CGH). The M-FISH analysis of this line did not reveal the enh(2)(q23→q31) because the chromosome der(2) appears as a single stained chromosome whose length doesn't differ from normal chromosomes 2, with an exception made for the *BCR*/*ABL* fusion gene at the telomeric

edge (Fig. 24). In the same work, the CGH analysis of the cell line revealed the presence of an amplified region that was interpreted as the *BCR*/*ABL* fusion gene present at the end of 2q and part of the chromosome 2 sequence present in the M17.

Figure 24. M-FISH of a representative K562 cell (from Neumann et al., 2001).



The analysis by FISH of this cell line showed the presence of an undiscovered amplified region that is much bigger than the simple telomeric presence of the *BCR*/*ABL* gene. The FISH technique is still a useful tool to study at a molecular level the chromosome structure, giving a higher resolution for chromosome rearrangements.

MATERIAL AND METHODS

1. Human lymphocytes primary cultures

Phytohaemagglutinin-stimulated human lymphocytes from peripheral blood were cultured for 72 hours, in RPMI1640 medium supplemented with 10% fetal calf serum and 2 mM L-glutamine at 37°C in 5% CO_2. To obtain metaphase chromosomes expressing the aphidicolin-inducible fragile sites, aphidicolin (0.4 μM) and ethanol (0.5%) were added to cultures 24 hours before fixing. To induce the DAPI-inducible fragile sites, DAPI (50 μg/ml) was added to cultures 20 hours before fixing.

2. Cell lines

A panel of 21 cancer cell lines was used. The cancer cell lines were derived from T cell acute lymphoblastic leukemia (LOUCY, MOLT-14 and PEER), Burkitt's lymphoma (DAUDI and RAJI), chronic myelogenous leukemia (two lines K562 by us named in this work K562-1 and K562-2), acute monocytic leukemia (MV4-11), erythroleukemia (HEL), colon carcinomas (DLD1, HCT116, SW620, CaCo2, HT29, LoVo and SW48), ovary adenocarcinoma (A2780), prostate carcinoma (DU145), endometrium carcinoma (HEC1A), cervix cacinoma (HeLa) and liver carcinoma (Hep-G2).

The lines LOUCY, MOLT-14, PEER, DAUDI, RAJI, K562, MV4-11, HEL were grown in RPMI1640 with 2 mM L-glutamine, 10% fetal bovine serum at 37°C in 5% CO_2; the tumor cell lines A2780, DU145, DLD1, HCT116, SW620,

CaCo2, HT29, LoVo, SW48, HEC1A, HeLa and Hep-G2 were grown in D-MEM High glucose with 2 mM L-glutamine, 10% fetal bovine serum at 37°C in 5% CO_2.

3. Spreads preparation

Colchicine (10^{-4} M) was added to cell culture 1-2 hours before harvesting. After incubation cells were harvested and treated with hypotonic solution (KCl 0.075M) for 8' at 20-22 °C. Cells were than fixed in cold methanol/acetic acid (3:1) and dropped on clean slides. Air-dried slides were kept at 20-22°C for aging and stored at 4°C for no longer than 4-6 months.

4. Probes and in situ hybridization

4.1 Cellular culture of E. coli and DNA extraction

BAC and PAC clones were used as probes for FISH experiment (see table below). The clones were selected by NCBI database (http://www.ncbi.nlm.nih.gov) and Ensembl Human Genome database (http://www.ensembl.org). The BAC and PAC clones were gently supplied from Dr. Mariano Rocchi, University of Bari (Italy), and/or ordered from Children's Hospital Oakland Research Institute (Oakland, CA).

Each colony was grown in selective LB medium at 37°C. To extract the BAC/PAC DNA, the bacteria were suspended in 50 mM Tris-HCl pH 8.5, 10 mM EDTA ph8; than treated with 0.2 M NaOH, 1% SDS and neutralized with 7.5 M CH_3COONH_4. DNA was precipitated and then resuspended in a suitable volume of TE pH 8 and RNase 50 μg/ml.

The BAC and PAC clones used for the molecular characterization of the common fragile site FRA7B were: RPCI-11 449P15 (AC073957), RPCI-11 6A1 (AC006433), RPCI-11 151M24 (AC024028), RPCI-11 42B7 (AC017000), RPCI-11 425P5 (AC009412), RPCI-11696N1 (AC004861), RPCI-4 755G17 (AC004879), RPCI-11 507C1 (AC007321), RPCI-708O1 (AQ516891-AQ408541).

The six BAC clones from the RPCI-11 library used for the replication timing analysis were: 295M18 (AL445423), 351P24 (AL513363), 281B4 (AQ507541), 527A7 (AC008177), 551O2 (AC008178), 284E18 (AC008065). The two cosmids used as early and late replicating control probes, were: GG1 (Smith and Higgs, 1999) and Me2.5 (Morley et al., 1991).

To investigate the presence of amplification events, the following probes were used: for FRA1H and/or FRA1I, the two BAC clones RP11-385F5 (AL359921) and RP11-434B7 (AL838261) were used, the BAC clones RP11-158L8 (AC073465) and RP11-681B22 (AC093684) were used for FRA2G/FRA2H, and 42B7 and 755G17 for FRA7B and the telomeric region.

To define the distal and proximal breakpoints of the duplicated regions identified on the marker chromosomes der(2) of the two analyzed cancer cell lines K562, other BAC clones were selected: to define the distal (telomeric) breakpoint were used 335G13 (AC097500), 489J10 (AC013445), 363G3 (AC092638), 351E10 (AC134880), to define the proximal (centromeric) breakpoint in the K562-1 cell line were used 472M4 (AC017082), 341F20 (AC013406), 375H16 (AC016731), 364H22 (AC005036); for the K562-2 cell line were used 11M1 (AC011178), 11N16 (AC018932), 527A7 (AC008177), 551O2 (AC008178), 724O16 (AC093899), 193F16 (AC009953) and 227L6 (AC078883).

To investigate the telomeric regions of the chromosomes der(2) of the two K562 cell lines two PAC probes (RPCI-5 1132H12 and RPCI-5 835J22), two BAC probes [434O9 (AQ584543-AZ303188) and 351E10 (AC134880)], and a telomeric probe $(TTAGGG)_n$ were used.

To define the fragile sequence of FRA2H, a set of seven BAC clones was used: 397C2 (AC104076), 418N16 (AC010888), 315C24 (AC093639), 335G13 (AC097500), 489J10 (AC013445), 88L20 (AC012488) and 317J9 (AC079777).

The new common fragile site FRA2S was identified by FISH using the BAC probes 341F20 (AC013406), 375H16 (AC016731) and 364H22 (AC005036) on metaphase spreads from cells of peripheral blood of five different healthy individuals induced with aphidicolin to express fragility.

4.2. Telomeric Probe

$(TTAGGG)_n$ probe was obtained by PCR using the protocol proposed by Ijido et al. (1991). PCR is carried in the absence of template using primers $(TTAGGG)_5$ and $(CCCTAA)_5$. Staggered annealing of the primers provides a single strand template for extension by *Taq* polymerase. PCR reactions were performed in 100 μl volumes containing 50 mM KCl, 10 mM Tris pH 8.3, 1.5 mM $MgCl_2$, 200 μM each nucleotide, 0.1 μM of each primer and 2 U of Taq polymerase. Amplification consisted of first 10 cycles of 1 min each at 94°C, 30 sec at 55°C and 1 min at 72°C, followed by 30 cycles each 1 min at 94°C, 30 sec at 60°C, 90 sec at 72°C, and one final step of 5 min at 72°C.

4.3. Nick Translation (NT)

Each clone and telomeric PCR product was labeled with biotin-16-dUTP or Digoxigenin-11-dUTP (ROCHE) by nick translation in a ratio of 65% modified nucleotide and 35% dTTP. For each nick translation reaction mix 1 μg of DNA was labeled for 2 hours at 15°C in NT buffer (1M Tris pH 8, 1M $MgCl_2$, BSA 100 mg/ml), with dNTP mix, 0.1M β-Mercaptoethanol, 0.02 U DNase I, 10 U *E. coli* DNA polymerase I, in order to obtain labeled fragments in the range of 300-600 bp.

4.4. Fluorescent in situ hybridization

For each FISH, 200 ng of labeled probe pre-annealed with CotI DNA (Invitrogen) was used. Slides were denatured in 70% formamide, 2xSSC at 80°C with bio/dig labeled probe.

Hybridization was carried overnight at 37°C in a moist chamber. Post-hybridization washes were performed at 60°C in 1xSSC.

4.5. Detection of probes

Biotin-labeled probes were detected with Cy3-conjugated avidin, and digoxigenin-labeled probes were detected with FITC-conjugated antidigoxigenin antibody. Slides were counterstained with DAPI (4'-6-diamidino-2-phenylindole, SIGMA). The images were recorded using a ZEISS Axioscope epifluorescence microscope equipped with a CCD camera and then merged using the Adobe Photoshop Software.

5. BrdU labeling

To identify the S-phase nuclei by immunolabeling detection, the cell cultures were supplied with BrdU (10^{-5} M), 20 min before fixing. S phase BrdU-labelled nuclei were visualized by indirect immunofluorescence immediately after FISH. Slides were incubated with 1 μg/ml of anti-BrdU antibody (Calbiochem) in PBS, 5% fetal calf serum for 1 hour. Slides were than incubated for 30 min with 15 μg/ml FITC-conjugated antimouse IgG (Vector Labs). Epifluorescence images were captured using a ZEISS Axioscope microscope equipped with a CCD camera. For confocal imaging a LEICA TMIRE2 inverted confocal microscope was used. The images were merged using the Adobe Photoshop Software.

6. Sequence analysis of fragile regions

The AT content and the composition in DNA repeted elements was analyzed using the RepeatMasker program (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker/), a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences.

Matrix Attachment Regions (MARs) were identified using the MAR-Wiz program (http://www.genomecluster.secs.oakland.edu/marwiz/).

CpG islands [observed/expected ratio >50, length >200 base pairs] were identified using the CPGPLOT program (http://www.ebi.ac.uk/emboss/cpgplot).

High flexibility regions were identified using the TwistFlex program (http://margalit.huji.ac.il/) that measures the local variations in the DNA structure and evaluates DNA flexibility, expressed as fluctuations of the DNA twist angle. The analysis was performed in overlapping windows of 100 bp. Adjacent overlapping windows with twist angle-scores exceeding 13.7° are referred as 'flexibility peaks'. When the distance within the last base of one peak and the first base on the next peak is smaller than the window size, these two peaks are considered as one single 'unified peak'. TwistFlex can also identify 'clusters' of flexibility peaks, which are at least three flexibility peaks that cover a window < 5 kb (Zlotorynski et al., 2003).

Predicted secondary structures formation of the flexible regions was performed using M-fold program (http://www.mfold.rna.albany.edu/mfold).

REFERENCES

Andreassen PR, D'Andrea AD, Taniguchi T (2004). ATR couples FANCD2 monoubiquitination to the DNA-damage response. *Genes Dev, 18: 1958-1963.*

Anglana M, Apiou F, Bensimon A, Debatisse M (2003). Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell, 114: 385-394.*

Aqeilan RI, Trapasso F, Hussain S, Costinean S, Marshall D, et al. (2007). Targeted deletion of Wwox reveals a tumor suppressor function. *Proc Natl Acad Sci U.S.A., 104: 3949-3954.*

Arlt MF, Miller DE, Beer DG, Glover TW (2002). Molecular characterization of FRAXB and comparative common fragile site instability in cancer cells. *Genes Chromosome Cancer, 33: 82-92.*

Arlt MF, Casper AM, Glover TW (2003). Common fragile sites. *Cytogenet Genome Res, 100: 92-100.*

Arlt MF, Xu B, Durkin SG, Casper AM, Kastan MB, Glover TW (2004). BRCA1 is required for common fragile-site stability via its G2/M checkpoint function. *Mol Cell Biol, 24: 6701-6709.*

Arlt MF and Glover TW (2010). Inhibition of topoisomerase I prevents chromosome breakage at common fragile sites. *DNA Repair, 9: 678-689.*

Barnes LD, Garrison PN, Siprashvili Z, Guranowski A, Robinson AK, et al. (1996). *Fhit*, a putative tumor suppressor in humans, is a dinucleoside 5',5'''-P^1,P^3-triphosphate hydrolase. *Biochemistry, 35: 11529-11535.*

Bednarek A, Keck-Waggoner CL, Daniel RL, Laflin KJ, Bergsagel PL, et al. (2001). *WWOX*, the FRA16D gene, behaves as a suppressor of tumor growth. *Cancer Res, 61: 8068-8073.*

Bois P and Jeffreys AJ (1999). Minisatellite instability and germline mutation. *Cell Mol Life Sci, 55: 1636-1648.*

Boldog F, Gemmill RM, West J, Robinson M, Robinson L, Li E, Roche J, Todd S, Waggoner B, Lundstrom R, Jacobson J, Mullokandov MR, Klinger H, Drabkin HA (1997). Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B. *Hum Mol Genet, 6: 193-203.*

Bosco N, Pelliccia F, Rocchi A (2010). Characterization of FRA7B, a human common fragile site mapped at the 7p chromosome terminal region. *Cancer Genet Cytogenet, 202: 47-52.*

Calin GA, Sevignani C, Dimitru CD, Hyslop T, Noch E, et al. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U.S.A., 101: 2999-3004.*

Casper AM, Nghiem P, Arlt MF, Glover TW (2002). ATR regulates fragile site stability. *Cell, 111: 779-789.*

Casper AM, Durkin SG, Arlt MF, Glover TW (2004). Chromosomal instability at common fragile sites in Seckel syndrome. *Am J Hum Genet, 75: 654-660.*
Chan KL, Palmai-Pallag T, Ying S, Hickson ID (2009). Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nat Cell Biol, 11: 753-760.*

Chang NS (2002). A potential role of p53 and WOX1 in mitochondrial apoptosis. *Int J Mol Med, 9: 19-24. Review.*

Chang NS, Doherty J, Ensign A (2003). JNK1 physically interacts with WW domain-containing oxidoreductase (WOX1) and inhibits WOX1-mediated apoptosis. *J Biol Chem, 278: 9195-9202.*

Chang NS, Pratt N, Heath J, Schultz L, Sleve D, et al. (2001). Hyaluronidase induction of a WWdomain-containing oxidoreductase that enhances tumor necrosis factor cytotoxicity. *J Biol Chem, 276: 3361-3370.*

Chesi M, Bergsagel PL, Shonukan OO, Martelli ML, Brents LA, et al. (1998). Frequent dysregulation of the c-*maf* proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. *Blood, 91: 4457-4463.*

Ciullo M, Debily MA, Rozier L, Autiero M, Billault A, Mayau V, El Marhomy S, Guardiola J, Bernheim A, Coullin P, Piatier-Tonneau D, Debatisse M (2002). Initiation of the breakage-fusion-bridge mechanism through common fragile site activation in human breast cancer cells: the model of *PIP* gene duplication from a break at FRA7I. *Hum Mol Genet, 11: 2887-2894.*

Coquelle A, Pipiras E, Toledo F, Buttin G, Debatisse M (1997). Expression of fragile sites triggers intrachromosomal

mammalian gene amplification and sets boundaries to early amplicons. *Cell, 89: 215-225.*

Corbin S, Neilly ME, Espinosa R III, Davis EM, McKeithan TW, LeBeau MM (2002). Identification of unstable sequences within the common fragile site at 3p14.2: implications for the mechanism of deletions within fragile histidine triad gene/common fragile site at 3p14.2 in tumors. *Cancer Res, 62: 3477-3484.*

Curatolo A, Limongi ZM, Pelliccia F, Rocchi A (2007). Molecular characterization of the human common fragile site FRA1H. *Genes Chromosomes Cancer, 46: 487-493.*

Debatisse M, Coquelle A, Toledo F, Buttin G (1998). Gene amplification mechanism: the role of fragile sites. *Recent Results Cancer Res, 154: 216-226.*

Dekaban A (1965). Persisting clone with an abnormal chromosome in a woman prevously irradiated. *J Nucl Med, 6: 740-746.*

Dessen P, Knuutila S, Huret JL (2010). Chromosome. *Atlas Cytogenet Oncol Haematol, July 2004.* Updated 2010.

Durkin SG, Arlt MF, Howlett NG, Glover TW (2006). Depletion of CHK1, but not CHK2, induces chromosomal instability and breaks at common fragile sites. *Oncogene, 25: 4381-4389.*

Durkin SG and Glover TW (2007). Chromosome fagile sites. *Annu Rev Genet, 41: 169-192.*

El Achkar E, Gerbault-Seureau M, Muleris M, Deutrillaux B, Debatisse M (2005). Premature condensation induces breaks at the interface of early and late replicating chromosome bands

bearing common fragile sites. *Proc Natl Acad Sci U.S.A., 102: 18069-18074.*

Fang JM, Arlt MF, Burgess AC, Dagenais SL, Beer DG, Glover TW (2001). Translocation breakpoints in *FHIT* and FRA3B in both homologs of chromosome 3 in an esophageal adenocarcinoma. *Genes Chromosomes Cancer, 30: 292-298.*
Feichtinger W and Schmid M (1989). Increased frequencies of sister chromatid exchanges at common fragile sites (1)(q42) and (19)(q13). *Hum Genet, 83: 145-745.*

Freudenreich CH (2007). Chromosome fragility: molecular mechanisms and cellular consequences. *Frontiers in Bioscience, 12: 4911-4924.*

Gardiner K (1995). Human genome organization. *Curr Opin Genet Dev, 5: 315-322. Review.*

Glover TW, Berger C, Coyle J, Echo B (1984). DNA polymerase a inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum Genet, 67: 136-142.*

Glover TW and Stein CK (1987). Induction of sister chromatid exchange at common fragile sites. *Am J Hum Genet, 41: 882-890.*

Glover TW and Stein CK (1988). Chromosome breakage and recombination at fragile sites. *Am J Hum Genet, 43: 265-273.*

Glover TW (1998). Instability at chromosomal fragile sites. *Recent Results Cancer Res, 154: 185-199. Review.*

Glover TW, Arlt MF, Casper AM, Durkin SG (2005). Mechanism of common fragile site instability. *Hum Mol Genet, 14: 197-205. Review.*

Gu Y, Shen Y, Gibbs RA, Nelson DL (1996). Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nat Genet, 13: 109-113.*

Handt O, Sutherland GR, Richards RI (2000). Fragile sites and minisatellite repeat instability. *Mol Genet Metab, 70: 99-105. Review.*

Hansen RS, Canfield TK, Lamb MM, Gartler SM, Laird CD (1993). Association of fragile X syndrome with delayed replication of the FMR1 gene. *Cell, 73: 1403-1409.*

Harrison JC and Haber JE (2006). Surviving the breakup: the DNA damage checkpoint. *Annu Rev Genet, 40: 209-235.*

Hecht F and Glover TW (1984). Cancer chromosome breakpoints and common fragile sites induced by aphidicolin. *Cancer Genet Cytogenet, 13: 185-188.*

Hellman A, Rahat A, Scherer SW, Darvasi A, Tsui LC, Kerem B (2000). Replication delay along FRA7H, a common fragile site on human chromosome 7. *Mol Cell Biol, 20: 4420-4427.*

Hellman A, Zlotorynski E, Scherer SW, Cheung J, Vincent JB, Smith DI, Trakhtenbrot L, Kerem B (2002). A role for common fragile site induction in amplification of human oncogenes. *Cancer Cell, 1: 89-97.*

Hewett DR, Handt O, Hobson L, Mangeldorf M, Eyre HJ, et al. (1998). FRA10B structure reveals common elements in repeat expansion and chromosomal fragile sites genesis. *Mol Cell, 1: 773-781.*

Hirsch B (1991). Sister chromatid exchanges are preferentially induced at expressed and nonexpressed common fragile sites. *Hum Genet, 87: 302-306.*

Howlett NG, Taniguci T, Durkin SG, D'Andrea AD, Glover TW (2005). The Fanconi anemia pathway is required for the DNA replication stress response and for regulation of common fragile site stability. *Hum Mol Genet, 14: 693-701.*

Hsu YY and Wang YH (2002). Human fragile site FRA16B DNA excludes nucleosomes in the presence of distamycin. *J Biol Chem, 277: 17315-17319.*

Huebner K and Croce CM (2001). FRA3B and other common fragile sites: the weakest links. *Nat Rev Cancer, 1: 214-221.*

Huebner K and Croce CM (2003). Cancer and the FRA3B/FHIT fragile locus: it's a HIT. *Br J Cancer, 88: 1501-1506.*

Ijdo JW, Wells RA, Baldini A, Reeders ST (1991). Improved telomere detection using a telomere repeat probe $(TTAGGG)_n$ generated by PCR. *Nucleic Acid Research, 19: 4780.*

Jeppesen C and Nielsen PE (1989). Photofootprinting of drug-binding sites on DANN using diazo- and azido-9-aminoacridine derivatives. *Eur J Biochem, 182: 437-444.*

Jiang Y, Lucas I, Young DJ, Davis EM, Karrison T, Rest JS, Le Beau MM (2009). Common fragile sites are characterized by histone hypoacetylation. *Hum Mol Genet, 18: 4501-4512.*

Keck CL, Zimonjic DB, Yuan BZ, Thorgeisson SS, Popescu NC (1999). Nonrandom breakpoints of unbalanced chromosome

translocations in human hepatocellular carcinoma cell lines. *Cancer Genet Cytogenet, 111: 37-44.*

Kuo MT, Vyas RC, Jiang LX, Hittelman WN (1994). Chromosome breakage at a major fragile site associated with P-glycoprotein gene amplification in multidrug-resistant CHO cells. *Mol Cell Biol, 14: 5202-5211.*

Laird C, Jaffe E, Karpen G, Lamb M, Nelson R (1987). Fragile sites in human chromosomes as regions of late-replicating DNA. *Trends Genet, 3: 274-281.*

Le Beau MM, Rassool FV, Neilly ME, Espinosa R, Glover TW, Smith DI, McKeithan TW (1998). Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Hum Mol Genet, 7: 755-761.*

Limongi MZ, Pelliccia F, Rocchi A (2003). Characterization of the human common fragile site FRA2G. *Genomics, 81: 93-97.*

Limongi MZ, Curatolo A, Pelliccia F, Rocchi A (2005). Biallelic deletion and loss of expression analysis of genes at FRA2G common fragile site in tumor-derived cell lines. *Cancer Genet Cytogenet, 181: 181-186.*

Lozzio CB and Lozzio BB (1957). Human chronic myelogenous leukemia cell line with positive Philadelphia chromosome. *Blood, 45: 321-324.*

Magenis RE, Hecht F, Lovrien EW (1970). Heritable fragile site on chromosome 16: probable localization of haptoglobin locus in man. *Science, 179: 85-87.*

Matsukuma S and Utakoji T (1978). Asimmetric decondensation of the L cell heterochromatin by Hoechst 33258. *Exp Cell Res, 113: 453-455.*

Mattei MG, Philip N, Pellissier MC, Piquet C, Mattei JF, Giraud F (1987). A new rare fragile site on human chromosome 2. *Cytogenet Cell Genet, 46: 658.*

McClintok B (1951). Chromosome organization and gene expression. *In: Cold Spring Arbor Symp Quant Biol, 16: 13-47.*

Millikangas S, Himberg J, Böhling T, Nagy B, Hollmén J, Knuutila S (2006). DNA copy number amplification profiling of human neoplasms.. *Oncogenomics, 25: 7324-7332.*

Mishmar D, Rahat A, Scherer SW, Nyakatura G, Hinzmann B, Kohwi Y, Mandel-Gutfroind Y, Lee JR, Drescher B, Sas DE, Margalit H, Platzer M, Weiss A, Tsui LC, Rosenthal A, Kerem B (1998). Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by cloning the human simian virus 40 integration site. *Proc Natl Acad Sci U.S.A., 95: 8141-8146.*

Mishmar D, Mandel-Gutfroind Y, Margalit H, Rahat A, Kerem B (1999). Common fragile sites: G-band characteristics within an R-band. *Am J Hum Genet, 64: 908-910.*

Morley BJ, Abbott CA, Wood WG (1991). Regulation of human fetal and adult globin genes in mouse erythroleukemia cells. *Blood, 78: 1355-1363.*

Mrasek K, Schoder C, Teichmann AC, Behr K, Franze B, Wilhelm K, Blaurock N, Claussen U, Liehr T and Weise A (2010). Global screening and extended nomenclature for 230

aphidicolin-inducible fragile sites, including 61 yet unreported ones. *Int J Onc, 36: 929-940.*

Musio A, Montagna C, Mariani T, Tilenni M, Focarelli ML, et al. (2005). SMC1 involvement in fragile site expression. *Hum Mol Genet, 14: 525-533.*

Naumann S, Reutzel D, Speiker M, Decker CK (2001). Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, and comparative genomic hybridization. *Leukemia Res, 25: 313-322.*
Nghiem P, Park PK, Kim YS, Vaziri C, Schreiber SL (2001). ATR inhibition selectively sensitizes G1 checkpoint-deficient cells to lethal premature chromatin condensation. *Annu Rev Genet, 41: 169-192.*

Niimura Y and Gojobori T (2002). In silico chromosome staining: reconstruction of Giemsa bands from the whole human genome sequence. *Proc Natl Acad Sci U.S.A., 99: 797-802.*

Ohta M, Inoue H, Cotticelli MG, Kastury K, Baffa R, et al. (1996). The *FHIT* gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell, 84: 587-597.*

O'Hagan RC, Chang S, Maser RS, Mohan R, Artandi SE, Chin L, DePinho RA (2002). Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell, 2: 149-155.*

O'Keefe RT, Henderson SC, Spector DL (1992). Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-

specific alpha-satellite DNA sequences. *J Cell Biol, 116: 1095-1110.*

Paige AJW, Taylor KJ, Taylor C, Hillier SG, Farrington S, et al. (2001). *WWOX*: a candidate tumor suppressor gene involved in multiple tumor types. *Proc Natl Acad Sci U.S.A., 98: 11417-11422.*

Palakodeti A, Han Y, Jiang Y, Le Beau MM (2004). The role of late/slow replication of the FRA16D in common fragile site induction. *Genes Chromosomes Cancer, 39: 71-76.*

Paradee W, Jayasankar V, Mullins C, Wilke C, Glover TW, Smith DI (1988). Molecular characterization of the 3p14.2 constitutive fragile site. *Am J Hum Genet, 55: A115.*

Paradee W, Wilke CM, Wang L, Shridhar R, Mullins CM, Hoge A, Glover TW, Smith DI (1996). A 350-kb cosmid contig in 3p14.2 that crosses the t(3;8) hereditary renal cell carcinoma translocation breakpoint and 17 aphidicolin-induced FRA3B breakpoints. *Genomics, 35: 87-93.*

Pekarsky Y, Garrison PN, Palamarchuk A, Zanesi N, Ageilan RI, et al. (2004). Fhit is a physiological target of the protein kinase Src. *Proc Natl Acad Sci U.S.A., 101: 3775-3779.*

Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM (2003). Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol, 23: 1349-1357.*

Pelliccia F and Rocchi A (1986). DAPI-inducible common fragile sites. *Cytogenet Cell Genet, 42: 174-176.*

Pelliccia F and Rocchi A (1988). Synergistic effect of DAPI and thymidylate stress conditions on the induction of common fragile sites. *Cytogenet Cell Genet, 48: 51-54.*

Pelliccia F, Curatolo A, Limongi ZM, Bosco N, Rocchi A (2007). Transcriptional profiling of genes at the human common fragile site FRA1H in tumor-derived cell lines. *Cancer Genet Cytogenet, 178: 144-150.*

Pelliccia F, Bosco N, Curatolo A, Rocchi A (2008). Replication timing of two human common fragile sites: FRA1H and FRA2G. *Cytogenet Genome Res, 121: 196-200.*

Pelliccia F, Bosco N, Rocchi A (2010). Breakages at common fragile sites set boundaries of amplified regions in two leukemia cell lines K562 - Molecular characterization of FRA2H and localization of a new CFS FRA2S. *Cancer Letters, 299: 37-44.*

Popescu NC and DiPaolo JA (1989). Preferential sites for viral integration on mammalian genome. *Cancer Genet Cytogenet, 42: 157-171.*

Popovici C, Basset C, Bertucci F, Orsetti B, Adelaide J, et al. (2002). Reciprocal translocations in breast tumor cell lines: cloning of a t(3;20) that targets the *FHIT* gene. *Genes Chromosomes Cancer, 35: 204-218.*

Portugal J and Waring MJ (1988). Assignment of DNA binding sites for 4',6-diamidine-2-phenylindole and bisbenzimide (Hoechst 33258). A comparative footprinting study. *Biochim Biophys Acta, 949: 158-168.*

Prantera G, Di Castro M, Cipriani L, Rocchi A (1981). Inhibition of human chromosome condensation induced by DAPI as related to cell cycle. *Exp Cell Res, 135: 63-68.*

Ragland RL, Glynn MW, Arlt MF, Glover TW (2008). Stably transfected common fragile site sequences exhibit instability at ectopic sites. *Genes Chromosomes Cancer, 47: 860-872.*

Rassool FV, McKeithan TW, Neilly ME, van Melle E, Espinosa RD, Le Beau MM (1991). Preferential integration of marker DNA into the chromosomal fragile site at 3p14: an approach to cloning fragile sites. *Proc Natl Acad Sci U.S.A., 88: 6657-6661.*

Reshmi SC, Huang X, Schoppy DW, Black RC, Saunders WS, Smith DI, Gollin SM (2007). Relationship between FRA11F and 11q13 gene amplification in oral cancer. *Genes Chromosomes Cancer, 46: 143-154.*

Ruiz-Herrera A, Castresana J, Robinson TJ (2006). Is mammalian chromosomal evolution driven by regions of genome fragility? *Genome Biol, 7: R115.*

Sadoni N, Langer S, Fauth C, Bernardi G, Cremer T, et al. (1999). Nuclear organization of mammalian genomes: polar chromosomes territories build up functionally distinct higher order compartments. *J Cell Biol, 146: 1211-1226.*

Samadashwily GM, Raca G, Mirkin SM (1997). Trinucleotide repeats affect DNA replication in vivo. *Nat Genet, 17: 298-304.*

Schwartz M, Zlotorynski E, Goldberg M, Ozeri E, Rahat A, Kerem B (2005). Homologous recombination and nonhomologous end-joining repair pathways regulate fragile site stability. *Genes Dev, 19: 2715-2726.*

Schwartz M, Zlotorynski E, Kerem B (2006). The molecular basis of common and rare fragile sites. *Cancer Lett, 232: 13-26. Review.*

Selig S, Okumura K, Ward DC, Cedar H (1992). Delineation of DNA replication time zones by fluorescent in situ hybridization. *EMBO J, 11: 1217-1225.*

Shi Y, Zou M, Farid NR, Paterson MC (2000). Association of Fhit (fragile histidine triad) a candidate tumor suppressor gene, with the ubiquitin-conjugating enzime hUBC9. *Biochem, 352: 443-448.*

Shuster MJ, Han L, Le Beau MM, Davis E, Sawicki M, Lese CM, Park NH, Colicelli J, Gollin S (2000). A consistent pattern of *RINI* rearrangements in oral squamous cell carcinoma cell lines supports a breakage-fusion-bridge cycle model for 11q13 amplification. *Genes Chromosomes Cancer, 28: 153-163.*

Sinclair PB, Harrison CJ, Jarosova M, Foroni L (2005). Analysis of balanced rearrangements of chromosome 6 in acute leukemia: clustered breakpoints in q22-q23 and possible involvement of c-MYB in a new recurrent translocation, t(6;7)(q23;q32~36) . *Haematologica, 90: 602-611.*

Siprashvili Z, Sozzi G, Barnes LD, McCue P, Robinson AK, et al. (1997). Replacement of Fhit in cancer cells suppresses tumorigenicity. *Proc Natl Acad Sci U.S.A., 94: 13771-13776.*

Smit A (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev, 6: 657-663.*

Smith ZE and Higgs DR (1999). The pattern of replication at a human telomeric region (16p13.3): its relationship to chromosome structure and gene expression. *Hum Mol Genet, 8: 1373-1386.*

Smith DI, Zhu Y, McAvoy S, Khun R (2006). Common fragile sites, extremely large genes, neural development and cancer. *Cancer Letters, 232: 48-57.*

Smith PP, Friedman C, Bryant EM, McDougall JK (2007). Viral integration and fragile sites in human papillomavirus-immortalized human keratinocyte cell lines. *Genes Chromosomes Cancer, 5: 150-157.*

Subramanian PS, Nelson DL, Chinault AC (1996). Large domains of apparent delayed replication timing associated with triplet repeat expansion at FRAXA and FRAXE. *Am J Hum Genet, 59: 407-416.*

Sutherland GR and Hecht F (1985). Fragile sites on human chromosomes. *In: Oxford Monographs on Medical Genetics, Oxford University Press, Oxford, UK.*

Sutherland GR and Richards RI (1995). The molecular basis of fragile sites in human chromosomes. *Curr Opin Genet Dev, 5: 323-327.*

Sutherland GR, Parslow MI, Baker E (1985). New classes of common fragile sites induced by 5-azacytidine and bromodeoxyuridine. *Hum Genet, 69: 233-237.*

Sutherland GR (2003). Rare fragile sites. *Cytogenet Gen Res, 100: 77-84.*

Svetlova EY, Razin SV, Debatisse M (2001). Mammalian recombination hot spot in a DNA loop anchorage region: a model for the study of common fragile sites. *J Cell Biochem Suppl, 36: 170-178.*

Takebayashi SI, Manders EM, Kimura H, Taguchi H, Okumura K (2001). Mapping sites where replication initiates in mammalian cells using DNA fibers. *Exp Cell Res, 271: 263-268.*

Takebayashi SI, Sugimura K, Saito T, Sato C, Fukushima Y, Taguchi H, Okumura K (2005). Regulation of replication at the R/G chromosomal band boundary and pericentromeric heterochromatin of mammalian cells. *Exp Cell Res, 304: 162-174.*

Taniguchi T and D'Andrea AD (2002). The Fanconi anemia protein, FANCE, promotes the nuclear accumulation of FANCC. *Blood, 100: 2457-2462.*

The BAC Resource Consortium: integration of cytogenetic landmarks into the draft sequence of the human genome (2001). *Nature, 409: 953-958.*

Thorland EC, Myers SL, Persing DH, Sarkar G, McGovern RM, et al. (2000). Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. *Cancer Res, 60: 5916-5921.*

Toledo F, Coquelle A, Svetlova E, Debatisse M (2000). Enhanced flexibility and aphidicolin-induced DNA breaks near mammalian replication origins: implications for replicon mapping and chromosome fragility. *Nucl Acids Res, 28: 4805-4813.*

Tsukasaki K, Miller CW, Greenspun E, Eshaghian S, Kawabata H, Fujimoto T, et al. (2001). Mutations in the mitotic check point gene, MAD1L1, in human cancers. *Oncogene, 20: 3301-3305.*

Ueda R, Kohanbash G, Sasaki K, Fujita M, Zhu X, Kastenhuber ER, et al. (2009). Dicer-regulated microRNAs 222 and 339 promote resistance of cancer cells to cytotoxic T-lymphocytes by down-regulation of ICAM-1. *Proc Natl Acad Sci U.S.A., 106: 10746-10751.*

Wang L, Paradee C, Mullins C, Shridhar S, Rosati R, Wilke CM, Glover TW, Smith DI (1997). Aphidicolin-induced FRA3B breakpoints cluster in two distinct regions. *Genomics, 41: 485-488.*

Wang L, Darling J, Zhang JS, Huang H, Liu W and Smith DI (1999). Allele-specific late replication and fragility of the most active common fragile site, FRA3B . *Hum Mol Genet, 8: 431-437.*

Widrow RJ, Hansen RS, Kawame H, Gartler SM, Laird CD (1998). Very late DNA replication in the human cell cycle. *Proc Natl Acad Sci U.S.A., 95: 11246-11250.*

Wilke CM, Hall BK, Hoge A, Paradee W, Smith DI, Glover TW (1996). FRA3B extends over a broad region and contains a spontaneous HPV16 integration site: direct evidence for the coincidence of viral integration sites and fragile sites. *Hum Mol Genet, 5: 187-195.*

Woinarowski JM, Trevino AV, Rodriguez KA, Hardies SC, Benham CJ (2001). AT-rich islands in genomic DNA as a novel target for AT-specific DNA-reactive antitumor drugs. *J Biol Chem, 276: 40555-40566.*

Yu S, Pritchard E, Kremer E, Lynch M, Nancarrow J, Baker E, Holman K, Mulley JC, Warren ST, Schlessinger D, Sutherland GR, Richard RI. (1991). Fragile X genotype characterized by an unstable region of DNA.. *Science, 252: 1179-1181.*

Yu S, Mangelsdorf M, Hewett D, Hobson L, Baker E et al. (1997). Human chromosomal fragile site FRA16B is an amplified AT-rich minisatellite repeat. *Cell 88: 367-374.*

Yunis JJ and Soreng AL (1984). Constitutive fragile sites and cancer. *Science, 226: 1199-1204.*

Zanesi N, Fidanza V, Fong LY, Mancini R, Druck T, et al. (2001). The tumor spectrum in Fhit-deficient mice. *Proc Natl Acad Sci U.S.A., 98: 10250-10255.*

Zimonjic DB, Durkin ME, Keck-Waggoner CL, Park SW, Thorgeirsson SS, Popescu NC (2003). *SMAD5* gene expression, rearrangements, copy number, and amplification at fragile site FRA5C in human hepatocellular carcinoma. *Neoplasia, 5: 390-396.*

Zlotorynski E, Rabat A, Skaug J, Ben-Porat N, Ozeri E, Hershberg R, Levi A, Scherer SW, Margalit H, Kerem B (2003). Molecular basis for expression of common and rare fragile sites. *Mol Cell Biol, 23: 7143-7151.*

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Prof. Franca Pelliccia. Her guidance helped me in all the time of research and writing of this thesis. I warmly thank Prof. Angela Rocchi for her constructive comments and immense knowledge. I owe my most sincere gratitude to Prof. Francesca Degrassi for her detailed and constructive comments. I could not have imagined having a better tutor for my Ph.D study.

During this work I have collaborated with many colleagues for whom I have great regard, and I wish to extend my warmest thanks to all those who have helped me with my work. I would never forget the great time I spent with Silvia Baldari, Carmela Viscomi, Valentina Ubertini, Simona Graziano and Ilaria Passacantilli. Thanks girls!

I am indebted to my many student and colleagues for providing a stimulating and fun environment in which to learn and grow. I am especially grateful to Eleonora Breno, Romina Burla and Antonella Friscini.

Lastly, and most importantly, I wish to thank Angela Curatolo. The joy and enthusiasm she has for her research was contagious and motivational for me, even during tough times in the Ph.D. pursuit. To her I dedicate this thesis.

Finally I have to thank my family for the constant love and support, for instilling in me confidence and a drive for pursuing my PhD, even if they never actually understood what I do in the laboratory.