CHAPTER 5

Cloning the homologous block rearrangement breakpoints in Siamang chromosome 18

5.1 Introduction

In chapter 4, the human chromosome 22q clone fF45C1 was identified by FISH, which contains sequences homologous to those that spanned an evolutionary rearrangement breakpoint in Siamang chromosome 18. In order to analyse the underlying mechanism which caused the rearrangement, the homologous block junctions need to be analysed at the sequence level.

5.1.1. Strategies for Analysing Sequences at the Homologous Block Junctions

The aim of the work described in this chapter was to use two different approaches to map the homologous block junctions in Siamang chromosome 18 and isolate the gibbon sequences spanning the fusion points between chromosome 22-homologous and non-homologous regions.

The first approach was to use PCR to assay for Siamang STSs on chromosome 18, using primer pairs defining STSs at 1 kb intervals in fF45C1. Long-range PCR would be used to amplify overlapping “tile-paths” of sequences from the two gibbon chromosome 18 segments. This would be carried out by using alternate sense and anti-sense primers, so that the sense primer of one STS would be used for PCR with the anti-sense primer of an adjacent STS. The rationale for this was that if alternate primers primed successfully in the human then priming might also be successful between the same alternate primer pairs in the gibbon. Assuming that was the case it seemed reasonable to expect that a failure of priming in the gibbon might indicate a discontinuity in the homologous sequences and that the primers lie either side of the homologous block breakpoint. In that case the two homologous sequences of DNA will be too far apart for extension to take place.

A prerequisite for the PCR approach was the design and use of primers defining STSs from fosmid fF45C1.

The second approach was to use radiolabelled Siamang STSs to identify and isolate Siamang cosmid clones containing sequences spanning the homologous block junction points. That approach required the construction and screening of a Siamang genomic cosmid library from high molecular weight genomic DNA.

The cloning capacity of the cosmid system is only a fraction of that which can be achieved using the large-insert PAC and BAC systems. But due to the high efficiencies of ligation, packaging and transfection, a cosmid library is much easier to prepare than a PAC or BAC library, even from small quantities of starting genomic DNA. A cosmid cloning system was chosen to be the most appropriate for this study as the target regions were small and it was envisaged that a Siamang genomic library would have only limited use.

The cosmid cloning system exploits the ability of λ bacteriophage to introduce a DNA molecule into an *E. coli* host cell (Collins and Hohn, 1978). The cosmid vector is a plasmid incorporating a 12 bp λ *cos* signal sequence, which is required for packaging the DNA into the bacteriophage pre-head structure. Most of the phage genome is excluded from the cosmid because there is no need to produce infectious virions following transfection of the host cell. The viral particles containing linear recombinant DNA molecules are assembled *in vitro*. Under the correct conditions these particles “infect” the host and introduce their DNA contents. Within the cell, the linear DNA is circularised and maintained as an extra-chromosomal plasmid. The bacteriophage head exerts a size limitation on the system as it will not accommodate a DNA molecule larger than 52 kb or smaller than 38 kb. Therefore there is no requirement to size-select the genomic DNA for ligation.

The Lawrist 16 vector (Yokobata *et al.,* 1991) has dual *cos* sites (Bates and Swift, 1983). Thus the recombinant DNA molecule comprises a fragment of genomic DNA flanked by vector arms, each with its own *cos* site. Vector concatemer formation is minimal, and so the genomic DNA is dephosphorylated instead of the vector. As well as dual *cos* sites, Lawrist 16 also has a *Bam*HI site for cloning fragments obtained by partial *Mbo*I or *Sau*3AI digestion, an origin of replication that will function in the host cell and a Neo^R gene for kanamycin selection

in transfected *E. coli*. These features are accommodated in only 8 kb of DNA; thus the cloned genomic DNA fragment can be as large as 44 kb.

An *E. coli* host strain that is suitable for the propagation of cosmid libraries has disabled recombination and restriction systems and also can be infected at high efficiency by λ phage. An *E. coli* strain with the appropriate genotypic features, which has been widely used for cosmid library propagation is DH5αMCR.

For libraries with very limited usage, the primary transfectants can be plated directly onto membranes, grown overnight, then used for preparation of replica membranes. The master membrane is stored at -70°C in the presence of glycerol, and positive clones are isolated from this membrane following hybridisation to its replica.

RESULTS

5.2 PCR analysis of Siamang chromosome 18

Fosmid fF45C1 is 45.681 kb in length and has now been fully sequenced at the Sanger Institute. At the time of this study, the clone had been through the sequencing pipeline but, due to secondary structures causing problems with the process, it was sequenced in two separate pieces A and B (Stephen Dodsworth, personal communication). Strand A was 6.355 kb and strand B was 38.926 kb in length. Based on the size of the insert and the two sequences, the gap between the two strands was estimated to be approximately 400 bp in length (Stephen Dodsworth, personal communication). Fosmid fF45C1 contains both Alu and LINE repeats, as well as a Colony Stimulating Factor (CSF2RB) gene and CSF-like pseudogene.

5.2.1 STS design and verification

Forty STSs with an average size of 141 bp and spaced at intervals of approximately 1 kb were identified (Sarah Hunt, Human Genetics Informatics, Sanger Institute) from the known sequence of strand A (A1 to A6) and strand B (B1 to B38) of fF45C1 (the sequences of the primers, plus the size of each STS are listed in Appendix I). The intervals between B16 and B18, B23 and B25, B25 and B27, and B28 and B30 were approximately 2 kb. For PCR assays of the STSs, primer pairs ("sense" and "anti-sense") were designed (Sarah Hunt, Human Genetics Informatics, Sanger Institute). Due to the repetitive nature of some of the fF45C1 strand B sequence, STSs B17, B24, B26 and B29 could not be designed. However, primer pairs defining 40 STSs were considered sufficient to carry out the first round of PCR analysis of the regions of Siamang chromosome 18 homologous to fF45C1.

The STSs were tested for their specificity by PCR on human genomic DNA, human chromosome 22-hybrid DNA and hamster genomic DNA. The appropriate reaction conditions were established to ensure that each assay amplified the expected size product prior to their use in characterising Siamang chromosome 18.

All 40 primer pairs amplified the predicted sized product from human genomic and human chromosome 22-hybrid DNA.

5.2.2 STS-PCR mapping of Siamang DNA

The STSs were tested for their presence on Siamang genomic DNA by PCR. Thirty-six of the primer pairs (90%) generated a specific product. Nine of the 36 successful primers (A1, B5, B11, B19, B20, B33, B35, B36 and B37) generated a specific product, as well as at least one other product. Five of the primer pairs (A3, B3, B5, B16 and B19) generated a product, which was a different size to that generated from human DNA, suggesting that a degree of sequence divergence had taken place. The STS-PCR results are summarised in Figure 5.1. The annealing temperatures most suited to each primer pair are tabulated below (Table 5.1).

Table 5.1 (next page) Optimal annealing temperature for each primer pair for human chromosome 22 and gibbon genomic DNA as PCR template. The optimal annealing temperature was selected as that which generated a single product. In the cases where more than one product was generated at all three temperatures, the optimal annealing temperature was chosen that gave the most specific product.

O.A.T. – optimal annealing temperature (˚C)

nsp – no specific product

[a] More than one product generated from human genomic DNA

[b] More than one product generated from human chromosome 22 hybrid DNA

[c] More than one product generated from Siamang genomic DNA

[d] Product from gibbon DNA is a different size to that from human DNA

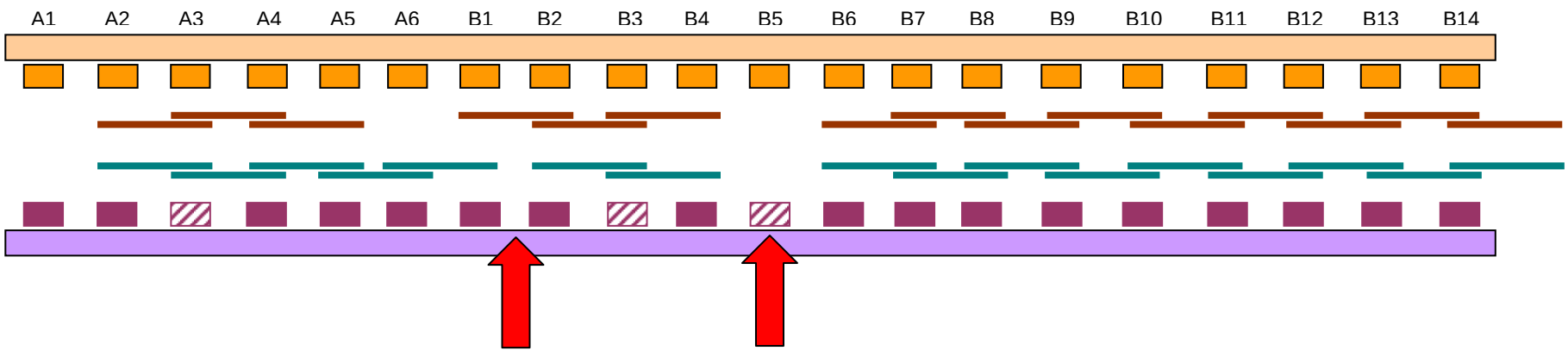
Primer Pair	O.A.T. Human (˚C)	O.A.T. Gibbon (˚C)	Primer Pair	O.A.T. Human (˚C)	O.A.T. Gibbon (˚C)
A1	65[a]	55[c]	B15	65	65
A2	60	65	B16	65[a]	65[d>]
A3	65	65[d>]	B18	65	nsp
A4	60	60	B19	65	65[c,d<]
A5	65	60	B20	65	60[c]
A6	65	65	B21	65	nsp
B1	65	65	B22	65[a,b]	nsp
B2	65	65	B23	65	nsp
B3	65	65[d<]	B25	65	65
B4	60	60	B27	60	60
B5	65	60[c,d<]	B28	60	60
B6	65	65	B30	60	60
B7	60	60	B31	60	60
B8	65	65	B32	60	60
B9	65	65	B33	60	60[c]
B10	60	60	B34	60	60
B11	65	65[c]	B35	60	60[c]
B12	60	60	B36	65[a,b]	65[c]
B13	65	65	B37	65	65[c]
B14	65	65	B38	65	65

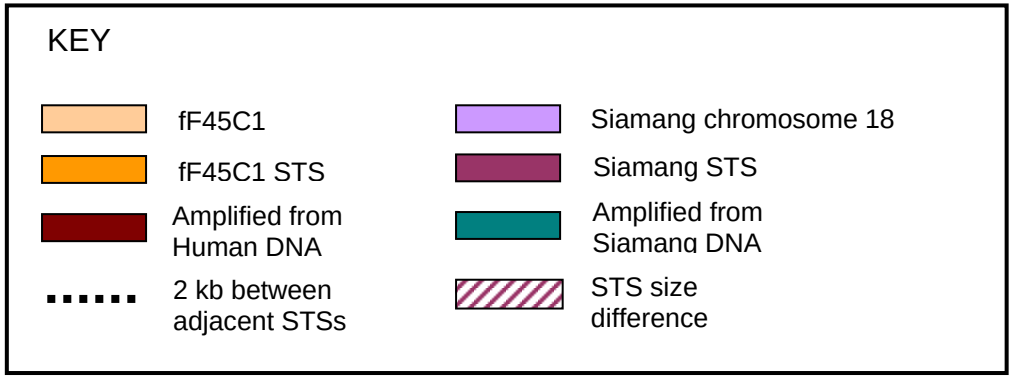
5.2.3 Long-range PCR assays of Siamang DNA

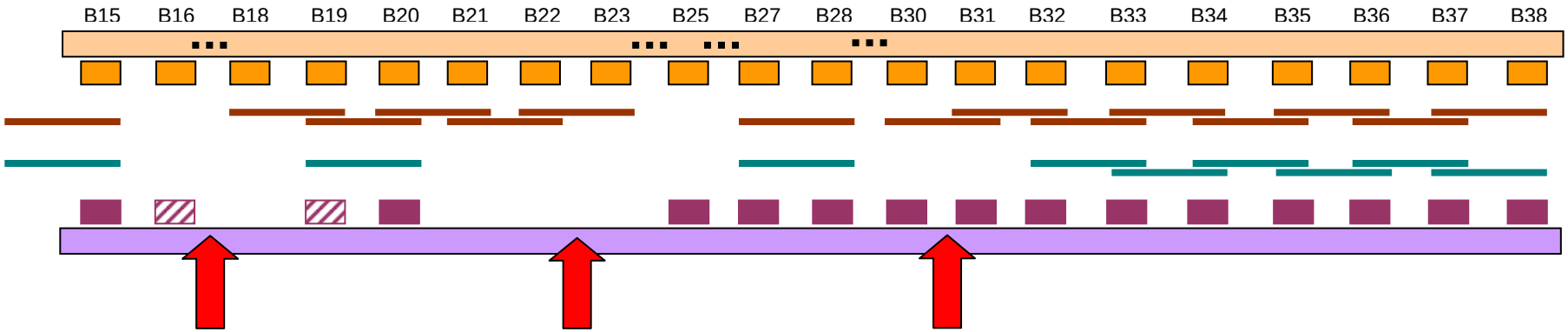
Based on the results from the STS analysis of Siamang DNA, alternate sense and anti-sense primers from fF45C1 STSs were used for longer-range PCR assays to amplify overlapping sequences from the two Siamang chromosome 18 segments. Individual oligonucleotides were mixed to collectively define overlapping sequences. For example, A1 sense was mixed with A2 anti-sense (A1 + A2), A3 sense with A4 anti-sense (A3 + A4), and so on.

For one reaction (A6 + B1) the expected PCR product size in human was approximately 1.4 kb. For four reactions (B16 + B18, B23 + B25, B25 + B27 and B28 + B30), the expected PCR product sizes were approximately 2 kb. For all other reactions, the expected PCR product sizes in human were approximately 1 kb. Long-range PCR assays were carried out on high molecular weight human and Siamang genomic DNA. The results of the long-range PCR assays are summarised in Figure 5.1 and presented as STS maps of human chromosome 22 clone fF45C1 and part of Siamang chromosome 18.

Figure 5.1 (next page) STS-PCR and long-range PCR map of fF45C1 and Siamang chromosome 18. The fF45C1 and Siamang chromosome 18 STSs are indicated by orange and purple boxes, respectively. The successful long-range PCR amplification between an alternate primer pair is indicated by a brown bar beneath two STSs for human DNA and by a green bar above two STSs for Siamang DNA. Stretches of gibbon sequence which failed to amplify are indicated by red arrows.


A1
A2
A3
A4
A5
A6
B1
B2
B3
B4
B5
B6
B7
B8
B9
B10
B11
B12
B13
B14


KEY
fF45C1
fF45C1 STS
Amplified from Human DNA
2 kb between adjacent STSs
Siamang chromosome 18
Siamang STS
Amplified from Siamang DNA
STS size difference


B15
B16
B18
B19
B20
B21
B22
B23
B25
B27
B28
B30
B31
B32
B33
B34
B35
B36
B37
B38

Twenty-nine of the 39 alternate primer pairs generated product of the expected size from human genomic DNA, and twenty-four from Siamang DNA. Two of the alternate primer pairs (A5 + A6 and A6 + B1) generated PCR product from Siamang DNA but not from human under long-range conditions. It might be that the sequence-feature in human that caused the original problem with sequencing fF45C1 also caused a problem in PCR but is not found in the gibbon sequence at that point. Alternatively, it could be that the PCR failures might be due to sequence divergence between the human and gibbon genomes.

The rationale for this long-range PCR analysis was that if an STS failed to amplify in the gibbon it might be implicated in the rearrangement point. But there were in fact five stretches of gibbon sequence, ranging in size from approximately 1 to 7 kb, which failed to amplify, listing below:

1. The expected size product from human DNA was generated by B1 + B2 primers, but no product was generated from gibbon DNA.
2. No product was generated from human or gibbon DNA by B4 + B5 or B5 + B6 primers.
3. No product was generated from human DNA by B15 + B16 or B16 + B18 primers. No product was generated from gibbon DNA by B15 + B16, B16 + B18 or B18 + B19 primers.
4. No product was generated from human DNA by B23 + B25 or B25 + B27. No product was generated from gibbon DNA from B20 + B21, B21 + B22, B22 + B23, B23 + B25 or B25 + B27.
5. No product was generated from human DNA by B28 + B30. No product was generated from gibbon DNA by B28 + B30, B30 + B31 or B31 + B32.

The individual primer pairs defining STSs B1, B2, B4, B5, B6, B15, B16, B19, B20, B25, B27, B28, B30, B31, and B32 generated the expected size product from Siamang DNA under normal PCR conditions.

Although the rationale was that a region, which fails to amplify in the gibbon, might be implicated in the rearrangement point, a region which fails in both human and gibbon could also be implicated. The only regions that can be excluded as the breakpoint are those which amplify the long product in the gibbon. Therefore, from the results of the long-range PCR analysis of Siamang DNA it was only possible to exclude 24 kb of fF45C1 sequence.

Because the long-range PCR approach did not provide a definitive answer in locating the position of the breakpoint in Siamang chromosome 18, an alternative approach was required, which was to generate a genomic gibbon library and screen for clones containing sequences spanning the fusion points.

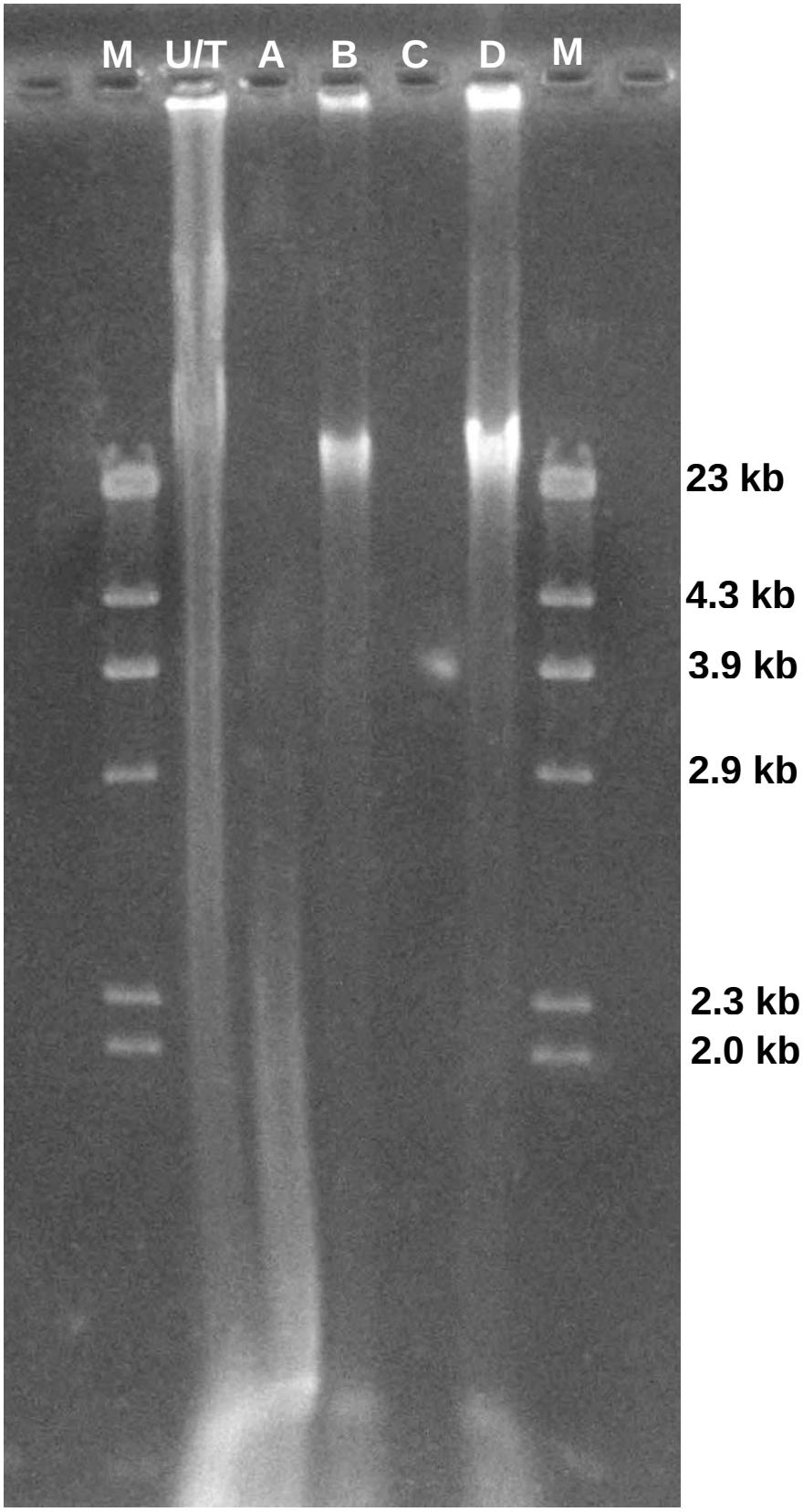
5.3 Construction of Siamang genomic cosmid library

5.3.1 Partial digestion of high molecular weight Siamang DNA

High molecular weight Siamang gibbon genomic DNA was extracted from cultured lymphoblastoid cells, and the OD_{260}/OD_{280} value was 1.875 indicating that there was no significant protein contamination in the DNA.

To assess its suitability for a competitive restriction digestion (see below), the high molecular weight DNA was subjected to restriction endonuclease digestion with *Mbo*I, both before and after methylation treatment with *dam* Methylase. The enzymes were included in excess to ensure that the reaction went to completion and that no *Mbo*I restriction sites remained which had not been cut or methylated. After gel electrophoresis it was deduced that the Siamang DNA had behaved similarly to human DNA. It was clear that *Mbo*I cut as predicted and that *dam* Methylase methylated and, thus, protected the Siamang DNA from digestion (Figure 5.2).

Figure 5.2 (next page) Photograph of 0.7% agarose gel of electrophoresed restriction digestion products. M= marker consisting of *Hind*III digested λ DNA (marker band sizes are indicated). U/T= untreated starting DNA. A= DNA treated with *Mbo*I. B= untreated control DNA. C= *Mbo*I enzyme only (no DNA). D= DNA treated with *dam* Methylase and *Mbo*I.


M
U/T
A
B
C
D
M
23 kb
4.3 kb
3.9 kb
2.9 kb
2.3 kb
2.0 kb

The DNA was cut by *Mbo*I to generate fragment sizes up to 4 kb. The DNA without enzyme resolved into a band, greater than 23 kb in size, as did the methylated DNA. Hoheisel *et al.* (1989) state that 6.25 units of *Mbo* I digests 1 µg of human genomic DNA to completion i.e. one cut in every 400 to 500 bp. Based on the current knowledge of human DNA, it was assumed that the Siamang DNA also has a *Mbo*I site approximately every 400 bases. In order to generate DNA fragments of sufficient size to clone in a cosmid library, DNA fragments of 40 to 50 kb were required. Therefore, it was necessary to digest the Siamang DNA at every 100th site to generate appropriately sized products.

5.3.1.1 Competitive digestion

The first approach used to digest the DNA was the competing *Mbo*I/*dam* methylase method described by Hoheisel *et al.* (1989). This was tried because competitive digestion was reported to be more controllable than methods relying on the use of limiting time or enzyme concentration. The technique was being used routinely in-house for the construction of flow-sorted chromosome-specific libraries (Ross and Langford, 1997).

A range of digestion and methylation conditions was tested, involving different unit ratios of *Mbo*I to *dam* Methylase (1:150, 1:75 and 1:300) and containing approximately 150 ng of DNA. The competing reactions of cleavage and methylation ran to completion. Using those conditions it was found that the Siamang genomic DNA cut too readily and fragment sizes were actually quite difficult to control. Even the products of the reaction containing the greatest ratio of *Mbo*I to *dam* Methylase (1:300) ranged in size from 2.3 to 23 kb, and were over-digested to consider for cloning.

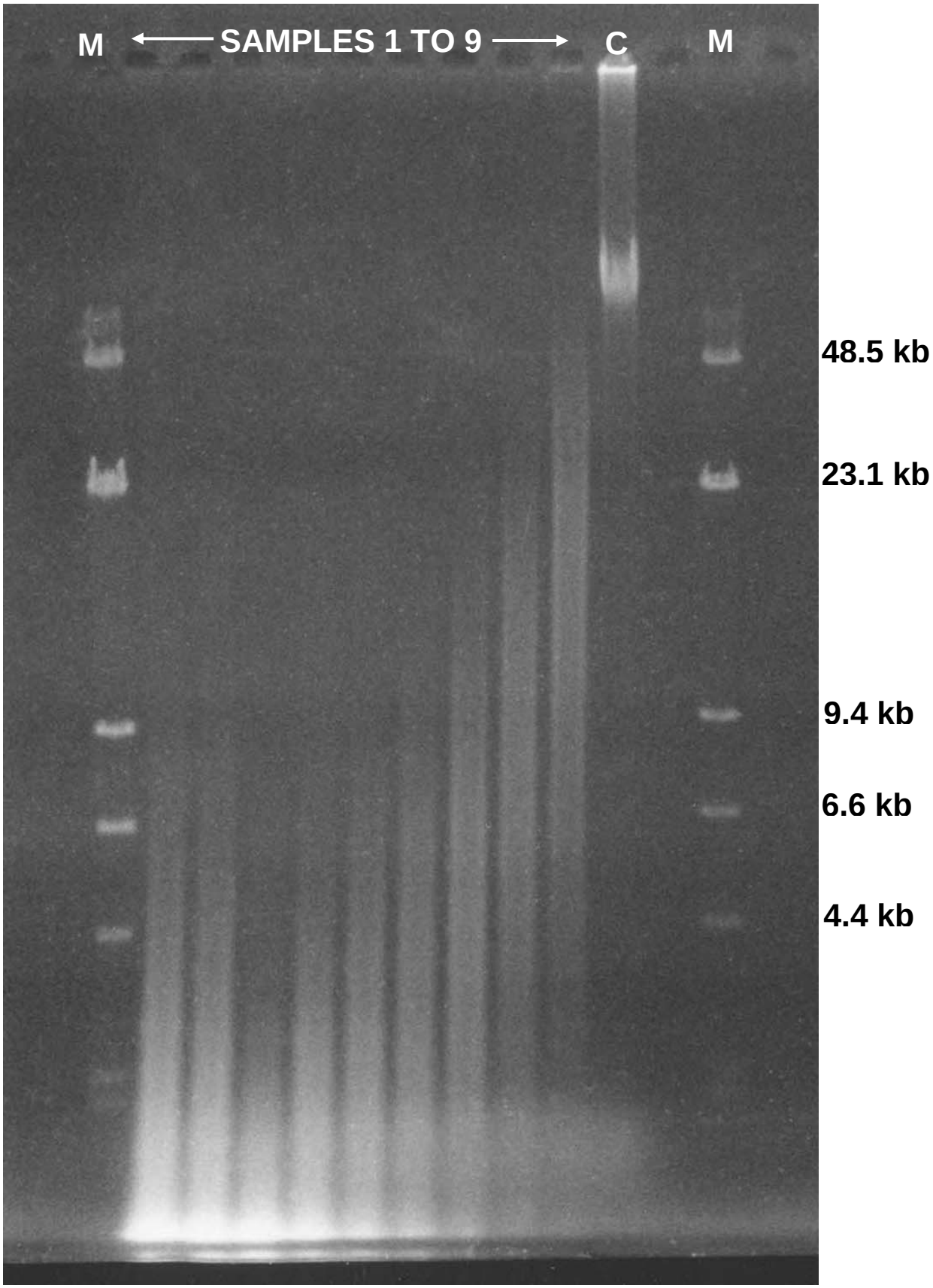
The volume of *dam* Methylase in the 1:300 unit ratio reaction (9.4 µl) was approximately one fifth of the total reaction volume (50 µl). The combination of 0.25 U of *Mbo*I with 75 U of *dam* Methylase resulted in a 10.4% v/v final concentration of glycerol. In order for the restriction enzymes to work effectively it was important to include less than or equal to 10% v/v glycerol in the reaction mix. It would have been impractical to set up reactions using higher ratios of *Mbo*I to *dam* Methylase without further increasing the proportion of glycerol. Therefore, it was decided not

to attempt further competitive reactions at higher ratios. Instead, a limiting enzyme approach was used by titrating *Mbo*I. This is described in the following section.

5.3.1.2 Limiting enzyme

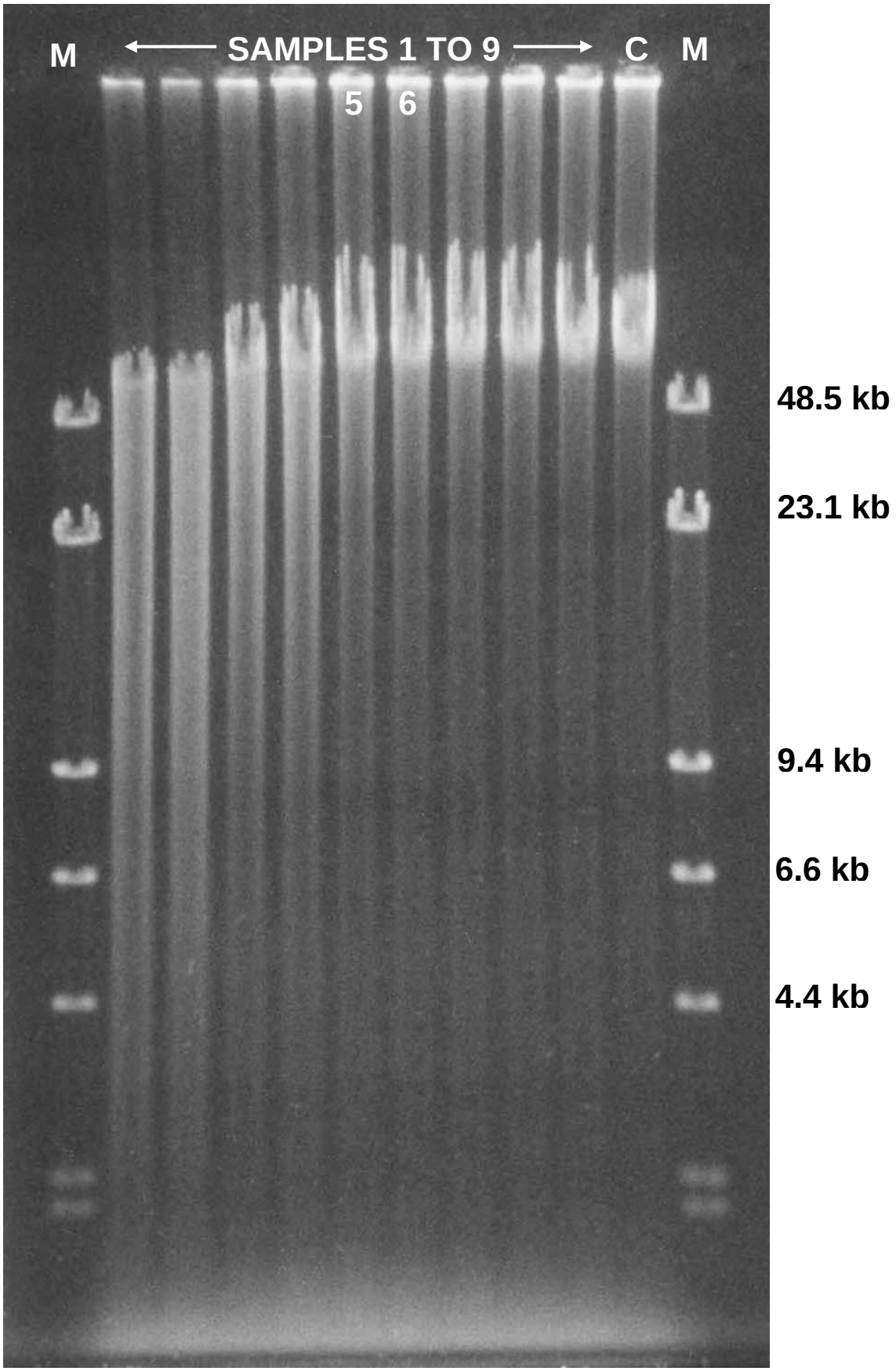
The second approach to digest the DNA was to use a dilution series of the *Mbo* I restriction enzyme alone for a fixed length of time (1 hr). Initially, nine partial digestion reactions were carried out involving a titration of *Mbo* I from 2 U down to 0.0078125 U, but in all of those reactions the DNA was over digested and ran ahead of the 23.1 kb size marker (Figure 5.3).

Figure 5.3 (next page) 0.3% agarose gel of electrophoresed restriction digestion products. M= marker consisting of *Hind*III digested λ DNA (marker band sizes are indicated). Samples as indicated. C= control lane, untreated DNA.


M
SAMPLES 1 TO 9
C
M
48.5 kb
23.1 kb
9.4 kb
6.6 kb
4.4 kb

A further nine reactions were carried out titrating *Mbo*I from 0.05 U down to 0.0001953125 U. Reactions 5 and 6 (0.003125 U and 0.0015625 U of *Mbo* I, respectively) produced optimal sized Siamang DNA restriction fragments for use in library construction (Figure 5.4). The optimal restriction reactions showed clear digestion, as in lanes 5 and 6, but with no evidence of DNA running ahead of the 23.1 kb λ *Hin*d III fragment.

Figure 5.4 (next page) 0.3% agarose gel of electrophoresed restriction digestion products. M= marker consisting of Hind*III digested λ DNA (marker band sizes are indicated). Samples as indicated. C= control lane, untreated DNA.*


M
SAMPLES 1 TO 9
C
M
5
6
48.5 kb
23.1 kb
9.4 kb
6.6 kb
4.4 kb

5.3.2 Ligation and packaging

Four reactions (ligation 1 to 4) were carried out to ligate the partially digested Siamang DNA to linearised Lawrist 16 vector arms. DNA from partial digestion reaction 6 (see *5.3.1.2*) was used for ligation 1 and 2, and DNA from digestion reaction 5 was used for ligation 3 and 4. In order to increase the chance of a successful out-come, each ligation reaction used a different nanogram ratio of vector arms to DNA. The ratios of arms to DNA for Ligation reactions 1 to 4 were 1:1, 8:1, 1.8:1 and 6:1, respectively.

Nine packaging reactions (HSY 1 to 9) of the ligated DNA were subsequently carried out to package the ligations into infective particles ready for plating on *E. coli*. 400 ng of DNA from ligation 1 was used for each packaging reaction HSY1 and HSY2. 63.3 ng of DNA from ligation 2 was used for HSY3, HSY4 and HSY5. 50 ng of DNA from ligation 3 was used for HSY6, HSY7 and HSY9 and 50 ng of DNA from ligation 4 was used for HSY8. In order to analyse the viability of commercially available packaging kits three different kits were used for the reactions. Gigapack Gold II was used for HSY1, HSY2, HSY3, HSY4, HSY5 and HSY6. Gigapack Gold IIXL was used for HSY7 and Gigapack Gold III was used for HSY8 and HSY9. (All three packaging kits were purchased from Stratagene).

5.3.3 Assessing the titre of the library

The best measure of the success of a library is determined by the genome coverage value. This requires the calculation of the total number of recombinant clones and the average clone insert size. The efficiency of cosmid library preparation, expressed as colony forming units (cfu) per µg of starting genomic DNA, is expected to lie between $5x10^5$ and $1x10^7$ for human DNA. The vector-arms self-ligation test (Section 2.11.3) gives a measure of the expected background of non-recombinants in the experiment. These non-recombinants can arise from vector concatemers caused by a failure to completely CIAP-modify the Lawrist 16 *Sca*I restriction sites.

Plating an aliquot of the packaged phage on *E. coli* gives the titre of the library and provides clones for an assessment of library quality (by FISH and restriction digestion). To test the titre of the nine packaging reactions, a sample of each was plated on *E. coli* and the results of the library titre assessments are summarised in Table 5.2.

Table 5.2 Assessment of the titres of each packaging reaction.

Packaging reaction	Reaction Volume (µl)	Ligated DNA (ng)	Colony Count	cfu[a] per µg DNA
HSY1	8	400	330	$5.4x10^5$
HSY2	8	400	260	$4.3x10^5$
HSY3	6.6	63.3	11	$1.1x10^5$
HSY4	6.6	63.3	9	$9.4x10^4$
HSY5	6.6	63.3	6	$6.2x10^4$
HSY6	8	50	67	$8.8x10^5$
HSY7	8	50	68	$8.9x10^5$
HSY8	4	50	32	$4.2x10^5$
HSY9	8	50	43	$5.7x10^5$

[a] cfu – colony forming units per µg of starting genomic DNA was calculated as follows: First, the total number of potential colonies for each packaging reaction was calculated by multiplying the number of colonies by 660 (the total available volume of each packaged extract). That value was then multiplied by 2.5 (for 400 ng ligated DNA), 15.798 (for 63.3 ng) or 20 (for 50 ng).

The background level of non-recombinants for ligation reactions 1 and 3 were defined by the number of colonies generated by the vector arms self-ligation reactions. The non-recombinants for ligations 2 and 4 were not assessed. The result obtained with the vector self-ligation should

indicate that less than 1% of clones are non-recombinant (Ross and Langford 1997). Less than 0.3% of clones from ligation reaction 1 and less than 0.1% of clones from ligation 3 were non-recombinant (data not shown).

The two packaging reactions HSY6 and HSY7 produced the highest titres (8.8×10^5 and 8.9×10^5 cfu's, respectively). They were both carried out in a reaction volume of 8 µl with 50 ng of ligated DNA (1.8:1 ratio of arms to DNA), using Gigapack Gold II (HSY6) and Gigapack Gold IIXL (HSY7) packaging kits. That result indicated that there was no advantage over the Gigapack Gold II in using the Gigapack Gold IIXL kit.

The next highest titre was produced by HSY9 (5.7×10^5 cfu's). That was also carried out in a reaction volume of 8 µl with 50 ng of DNA from the same ligation reaction (1.8:1 ratio) as for HSY6 and HSY7. The only difference, which may account for the lower titre, was the use of the Gigapack Gold III kit.

The packaging reactions HSY6 and HSY7 were approximately twice as efficient as reactions HSY1, HSY2 and HSY8 (5.4×10^5, 4.3×10^5 and 4.2×10^5 cfu's, respectively). The HSY1 and HSY2 reactions were also carried out in a volume of 8 µl, using the Gigapack Gold II kit, but a total of 400 ng of DNA ligated in a ratio of 1:1 was used, which may explain the lower titre.

The Gigapack Gold III kit was used for the HSY8 reaction and was carried out in a volume of only 4 µl, using 50 ng of DNA ligated in a ratio of 6:1.

Packaging reactions HSY3, HSY4 and HSY5 generated the lowest titres (1.1×10^5, 9.4×10^4 and 6.2×10^4 cfu's, respectively). Those reactions were carried out using the Gigapack Gold II kit, but in a reaction volume of 6.6 µl using 63.3 ng of DNA ligated in a ratio of 8:1.

It was concluded that the highest library titres were generated using 50 ng of DNA from a ligation reaction with a low ratio of vector arms to DNA (e.g. 1:1 or 1.8:1) and packaged in a volume of 8 µl using either Gigapack Gold II or IIXL kits.

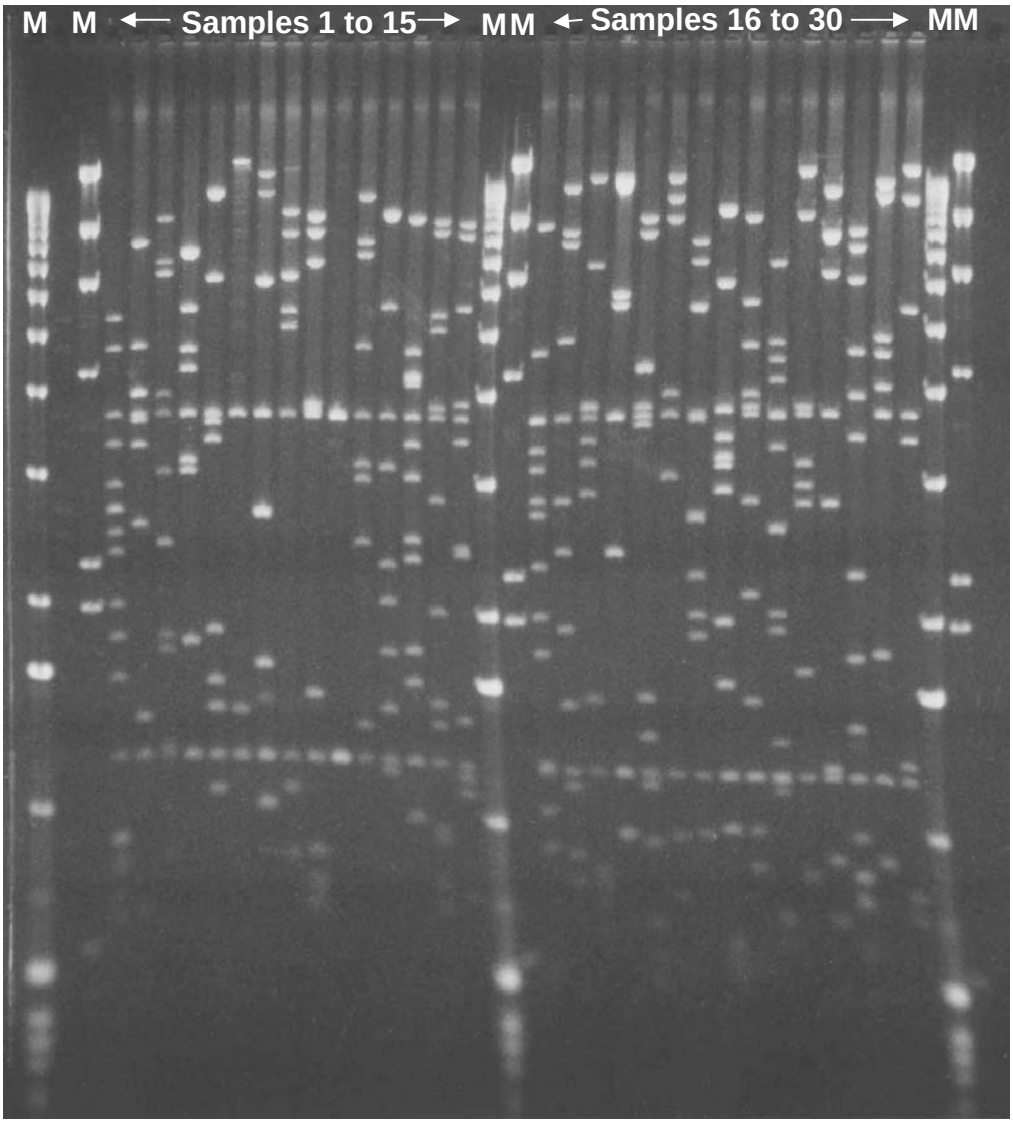
5.3.4 Assessing the integrity of Siamang genomic cosmid clones

5.3.4.1 EcoR I digestion to estimate insert size

The average insert size was estimated for thirty clones picked from the test plating for packaging reaction HSY1. Colonies (HSY1A1 to HSY1C6) were picked into LB freezing broth and individually grown up overnight. DNA was prepared by a standard alkaline-lysis procedure then digested with *EcoR* I. Samples of the cultures were stored frozen at –70˚C.

The products of the clone restriction digestion were analysed by electrophoresis together with size standards (1kb ladder and *Hin*d III digestion of λ DNA), on a 0.7% agarose gel. The gel was stained with 0.5 µg/ml Ethidium Bromide and the bands visualised under U.V. light (Figure 5.5).

Figure 5.5 (next page) 0.7% agarose gel with *Eco*RI digestion products for gibbon genomic cosmid clones. M= size marker. Samples as indicated.


M M
Samples 1 to 15
MM
Samples 16 to 30
MM

Due to the complexity of the gel it was difficult to manually estimate the total size of the inserts for each of the clones analysed. Further analysis of the clones was carried out by another *EcoR* I digestion followed by electrophoresis on a 0.7% agarose gel. Following staining with a 1:10,000 dilution of Vistra Green™ in TE, the gel was scanned automatically using a FluorImager scanner (Molecular Dynamics Inc). The scanner analysis software provided an estimate of the band sizes for each clone analysed (data not shown).

For Lawrist 16, insert sizes normally lie between 33 and 44 kb. One of the clones, (A10), had no insert and only vector bands were visible on the gel in the corresponding lane. The average estimated size of insert of the remaining 29 clones analysed was 37.1 kb, which is within the range expected.

Although the insert size figure of 37.1 kb was calculated from clones from a single library (HSY1), the same value was used to estimate the coverage of all the libraries, based on the assumption that HSY2 to HSY9 all contained similar sized inserts to HSY1. Coverage (C) was calculated from the equation:

C = [N x S]/G where

N =the total number of clones in the library

S = the average clone insert size, and

G =the size of the starting genome

The estimated genome coverage of each library is tabulated below (table 5.3).

Table 5.3 Estimated genome coverage of each library

Packaging reaction	Genome coverage
HSY1	2.7
HSY2	2.1
HSY3	0.09
HSY4	0.07
HSY5	0.05
HSY6	0.55
HSY7	0.56
HSY8	0.26
HSY9	0.35

Based on the assumption above, the total coverage of all the libraries together was calculated to be 6.73. However, as different genomic digests were used for the different libraries, the calculation was treated as an estimate. Library HSY7, which used the Gigapack Gold IIXL packaging system might be expected to have a higher average insert size, although this was not assessed.

The size of the starting genome determines the number of clones required for a given coverage, which in turn determines the probability of identifying a specific clone. The theoretical probabilities of finding at least one clone containing a given target sequence are 0.632, 0.993 and 0.999, respectively, for libraries with C values of 1, 5 and 7 (Mark Ross, personal communication). The overall Siamang genomic library coverage was 6.73, which provided a probability between 0.993 and 0.999 of identifying a clone from a specific point in the gibbon genome. It was concluded that

enough clones were present in the libraries to progress onto the screening part of the project (section 5.4).

5.3.4.2 FISH analysis to test clone specificity

FISH analysis of the cosmid DNA samples on metaphase chromosome spreads provides a measure of the frequency of chimaeric clones. DNA isolated from cosmid clones HSY1A1 to HSY1A10 was biotinylated by nick translation and hybridised to Siamang and human metaphase chromosomes. As would be expected, the probe for clone A10 gave no hybridisation signal on the Siamang chromosomes. One of the other clones hybridised to the end of every Siamang chromosome, but did not hybridise to any human chromosomes. It seems likely that it contains a heterochromatic/telomeric repeat, specific to the Siamang. The remaining eight clones hybridised with good, clear, single signals to both Siamang and human chromosomes.

5.4 Identification of Siamang cosmid clones spanning the homologous block junctions

5.4.1 Screening high-density filters

5.4.1.1 Plating high-density filters

The most suitable approach for screening the cosmid library was to hybridise radiolabelled gibbon STS probes to membrane filters carrying the DNA from lysed colonies. As a first attempt, it was decided to plate out and screen 2 x 10^5 colonies, which represented approximately 2.5 Siamang genome equivalents. The primary transfectants were randomly plated directly onto membranes on agar and grown overnight. The packaging reactions HSY1 and HSY2 were plated on *E. coli* to generate approximately 20,000 colonies on each of 10 master filters of 7.5 x 11 cm (SCHSY1 to SCHSY10). SCHSY1, 2, 3, 4 and 5 were generated from library HSY1, whereas SCHSY6, 7, 8, 9 and 10 were generated from library HSY2.

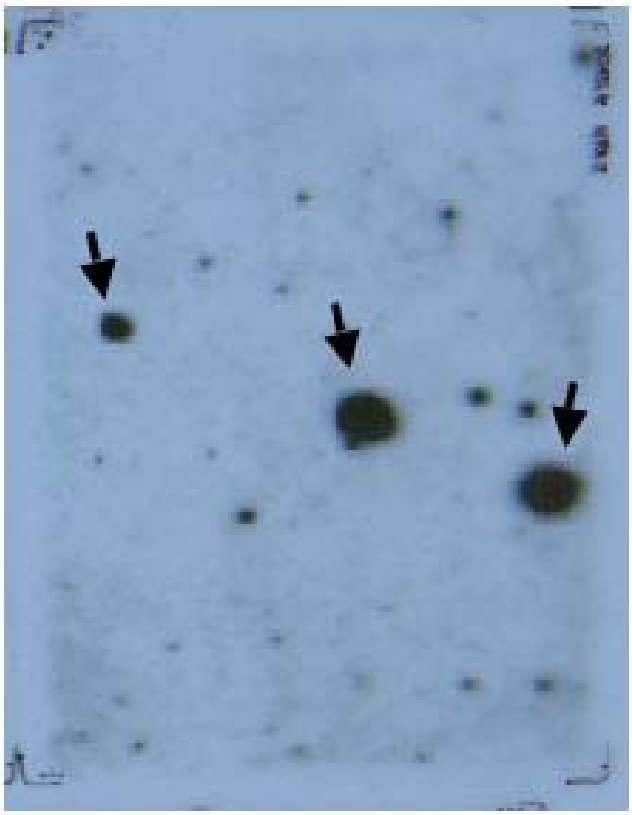
After the preparation of replica membranes, the master membrane was stored at -70°C in the presence of glycerol, so that clones could be isolated from this membrane following hybridisation to its replica. The replica filters, "SCHSYRep1" to "SCHSYRep10" were processed prior to screening.

5.4.1.2 High-density filter screening

The SCHSYRep filters were screened by hybridisation with pools of radiolabelled gibbon STSs. The PCR products were generated using the fF45C1 A4, B1, B10, B20 and B38 primer pairs on Siamang genomic DNA template. These primer pairs were chosen as they were distributed along the length of fF45C1 and spaced approximately 10 kb apart. The PCR products were excised from agarose gels then used as template in a second PCR containing a single radiolabelled dNTP (α-^{32}P dATP). The labelled products were pre-reassociated with gibbon genomic DNA, in order to compete out any repetitive sequences within the probes.

The probe pool was hybridised to the colony filters overnight at 65°C. Filters were then washed to remove any un-bound probe. After a 3.5-hour exposure, the autoradiograph showed a low background with numerous signals ranging in intensity. There were five very intense signals SCHSY4.1, SCHSY4.2, SCHSY5.1, SCHSY7.2 and SCHSY10.1. There were eighteen medium-intensity signals on SCHSY2, SCHSY4, SCHSY5, SCHSY6, SCHSY7, SCHSY8, SCHSY9 and SCHSY10. (see Figure 5.6 (left-hand panel) for an example filter). SCHSY1 and SCHSY3 had only numerous low-intensity signals.

Figure 5.6 (next page) From left to right, autoradiographs of high-density and low-density colony filters after probing with radiolabelled gibbon STS. In the left panel, strong signals are indicated with arrows. In the right panel, the signal from a colony, which was picked for subsequent analysis is indicated an arrow.



5.4.1.3 Colony verification by PCR

In an initial verification analysis, eleven “mixed colony” regions (SCHSY10.1 to SCHSY10.11), representing the full range of autoradiograph signal intensities, were picked from master filter SCHSY10, and streaked on LB agarose plates containing kanamycin. The entire areas demarcated by the autoradiograph signals (incorporating several colonies) were sampled from the master filter, to avoid the possibility of missing the positive colonies. After an overnight incubation, each mixed colony streak was sampled and analysed for the presence of gibbon STSs by colony PCR using fF45C1 A4, B1, B10, B20 and B38 primer pairs. Only SCHSY10.1 was positive containing A4 and B1 STSs. SCHSY10.1 was sampled from the region corresponding to the largest autoradiograph signal for filter SCHSYRep10.

Bearing the results of the initial PCR screen in mind, another twenty-one mixed colony regions were sampled and streaked from master filters SCHSY2, 4, 5, 6, 7, 8 and 9, representing the large and medium-sized autoradiograph signals. Each mixed colony streak was sampled and analysed by colony PCR as before.

After sampling for colony PCR, the mixed colony streaks were grown in liquid culture overnight prior to being frozen as “SCHSY mixed colony glycerols”. SCHSY4.1, 4.3, 5.1 and 7.2 contained STSs A4 and B1. SCHSY4.2 contained STSs B10 and B20. All the other cultures tested were negative.

5.4.2 Screening low-density filters

5.4.2.1 Estimating cell viability in SCHSY mixed colony glycerols

From a $1/10^6$ dilution in LB broth containing kanamycin, the concentration of viable cells for each SCHSY4.1, 4.2, 4.3, 5.1, 7.2, 8.4 and 10.1 mixed colony glycerol was estimated. The mixed colony dilutions were stored at 4˚C until used for plating out. The cell counts per ml for each mixed glycerol are tabulated below (Table 5.4).

Table 5.4 Cell counts per ml of each mixed colony glycerol.

SCHSY mixed colony glycerol	Cell count per ml	Vol. for 500 colonies[a] (µl)
4.1	$9.9x10^8$	505
4.2	$9.1x10^8$	550
4.3	$9.6x10^8$	521
5.1	$7.2x10^8$	694
7.2	$7.5x10^8$	667
8.4	$9.0x10^8$	555
10.1	$5.2x10^8$	961

[a] Volume of the mixed colony glycerol $1/10^6$ dilution (described above) containing 500 colonies

5.4.2.2 Plating low-density colonies

Approximately 500 colonies from each mixed colony $1/10^6$ dilution were spread on an aged LB agar plate containing kanamycin and incubated for 12 hours at 37˚C. Colony material was transferred from each plate onto a hybridisation membrane, which was processed immediately to prevent drying out. The colony plates were incubated for a further 2 to 3 hours at 37˚C to allow the colonies to re-grow before being stored at 4˚C.

5.4.2.3 Low-density filter screening

The low-density filters were screened by hybridisation with radiolabelled Siamang STSs as described above. SCHSY4.1, 4.3, 5.1, 7.2, 8.4 and 10.1 were screened with B1, and SCHSY4.2 was screened with B10.

After a 4.5-hour exposure, the autoradiograph showed a low background. There were numerous signals of similar intensities from filters 4.1 and 7.2. There were four high-intensity signals from 4.2 and only one low-intensity signal from 10.1. There were no signals from filters 4.3, 5.1 and 8.4. (Figure 5.6, see above)

5.4.2.4 Verifying colonies by PCR

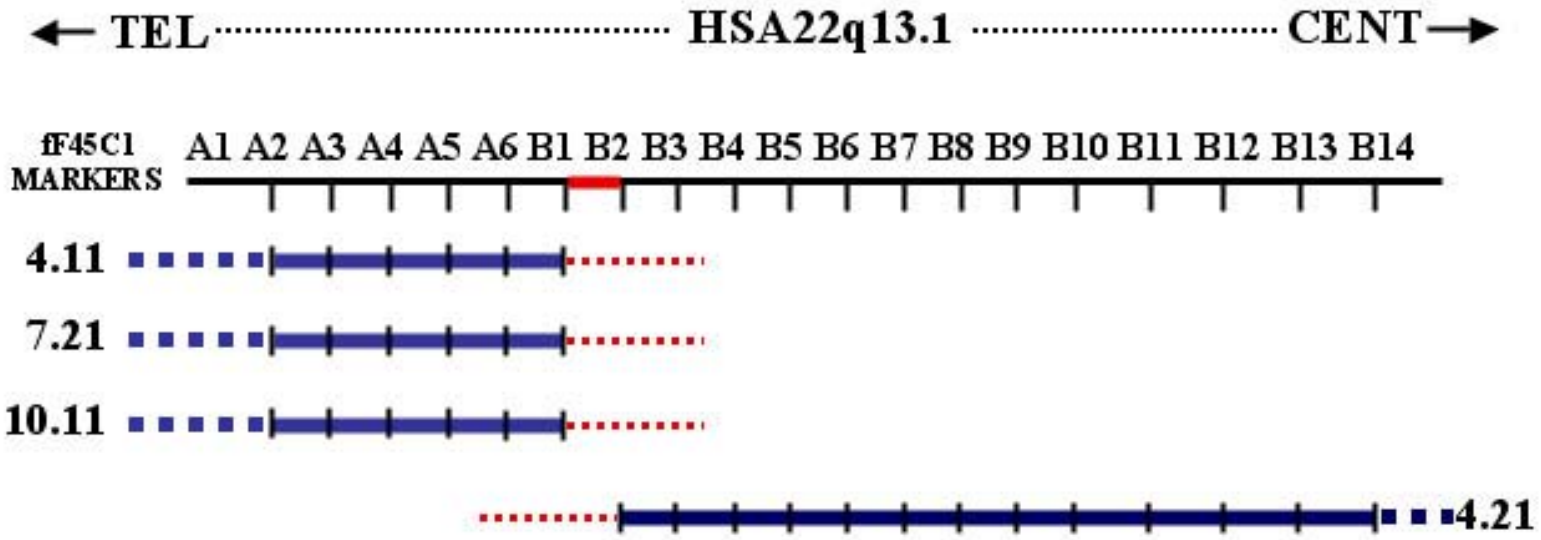
Single positive colonies (SCHSY4.11, 4.21, 7.21 and 10.11) were picked from the plates representing filters SCHSY4.1, 4.2, 7.2 and 10.1 and streaked to single colonies on fresh plates. One well-separated colony from each streak was picked and analysed for STS content by colony PCR using fF45C1 B1 and B10 primer pairs. SCHSY4.11, 7.21 and 10.11 contained the B1 STS but not the B10 STS. SCHSY4.21 contained the B10 STS, but not the B1 STS.

5.5 Construction of a cosmid map defining the homologous block rearrangement point

PCR analysis was carried out using fF45C1 primer pairs A2 to B38 to determine the STS marker content of each isolated cosmid and hence to identify overlaps between the clones. Clones SCHSY4.11, 7.21 and 10.11 contained STSs A2 to B1. Clone SCHSY4.21 contained STSs B2 to B20, inclusive, plus B25, B27 and B32. The four clones were organised into a map based on their regions of homology with human fF45C1 (summarised in Figure 5.7). From the results of the STS-PCR analysis of the four clones generated to this point it was possible to exclude the regions A2-B1 and B2-B32 from containing sequences spanning the breakpoint. The region B32-B38 had been previously excluded from containing breakpoint sequences based on the long-range PCR assays reported in section *5.2.3*.

Figure 5.7 (next page) STS map based on regions of homology of four gibbon cosmid clones with human chromosome 22 clone fF45C1. SCHSY4.11, 7.21 and 10.1 overlapped each other in the map as they all contained sequence homologous to the region in fF45C1 from STS A2 to B1. The region highlighted in red represents a section in fF45C1 not represented in the clones analysed.

Positive clones analysed for STS marker content and contig constructed



5.6 FISH analysis of breakpoint clones

Based on the results described above and the original FISH results of fF45C1, it was reasoned that DNA from the three clones SCHSY4.11, 7.21 and 10.11 would hybridise by FISH to Siamang 18p, and that DNA from SCHSY4.21 would hybridise to Siamang 18q.

DNA from the cosmids SCHSY4.11, 4.21, 7.21 and 10.11 was isolated and biotinylated by nick translation and hybridised to Siamang and human metaphase chromosomes in separate experiments. At least ten metaphase spreads were analysed for each hybridisation experiment. The DNA from SCHSY4.11, 7.21 and 10.11 hybridised to Siamang 18p. The DNA from SCHSY7.21 and 10.11 also hybridised to human 22q13.1. The DNA from SCHSY4.11 hybridised to human 22q13.1 and human 16p. These results are summarised in Table 5.5.

The DNA from SCHSY4.21 hybridised to Siamang 18q, human 22q13.1 and human 16p (see Table 5.5).

Table 5.5 Summary of FISH localisation of gibbon cosmids after hybridisation to gibbon and human metaphase chromosomes. A tick in a box indicates a positive hybridisation signal.

SCHSY cosmid	HSY 18p	HSY 18q	HSA 22q13.1	HSA 16p
4.11	3		3	3
7.21	3		3	
10.11	3		3	
4.21		3	3	3

These FISH results confirm that cosmid clones SCHSY4.11 and 4.21 contain gibbon sequences homologous to human 22q13.1 (4.21 proximal and 4.11 distal to the homologous block breakpoint) as well as containing sequences homologous to human 16p. The most likely explanation for these observations is that SCHSY4.11 contains the gibbon sequence, which spans the junction between the human chromosome 22- and 16-homologous blocks on Siamang 18p and SCHSY4.21 spans the junction on Siamang 18q.

The lack of signal on human chromosome 16p from clones SCHSY7.21 and 10.11 does not mean that they do not contain sequences that span the junction and don't contain sequences homologous to human 16p. They may well contain those sequences, but the stretch of DNA homologous to human chromosome 16p may be too short to generate a visible FISH signal.

Gibbon cosmid SCHSY4.11 contained STS markers A2 to B1 and it was FISH-mapped to HSY 18p, HSA 22q13.1 and HSA 16p. Cosmid SCHSY4.21 contained STS markers B2 to B20, B25, B27 and B32 and was FISH mapped to HSY 18q, HSA 22q13.1 and HSA 16p. In view of all the evidence, it was concluded that the evolutionary rearrangement breakpoint lies in material

homologous to human chromosome 22q13.1 within the 1 kb region defined by markers B1 and B2.

5.7 Discussion

This chapter describes the cloning and mapping of two homologous block evolutionary rearrangement junctions in Siamang chromosome 18.This involved the design of PCR primer pairs defining STSs in the human chromosome 22 fosmid fF45C1. Homologous STSs were mapped on Siamang chromosome 18 and five stretches of Siamang sequence failed to amplify by long-range PCR.

A Siamang genomic cosmid library of 6.73 genome equivalents was constructed from high molecular weight DNA. Part of the library was screened with radiolabelled Siamang STSs generated using fF45C1 PCR primers. The isolated clones were screened for STS marker content and organised into a map indicating homology with fF5C1. Two of the clones SCHSY4.11 and SCHSY4.21 were shown by FISH to contain sequences homologous to human 22q and human 16p. SCHSY4.11 was from Siamang chromosome 18p and 4.21 was from chromosome 18q. The evidence of the FISH results coupled with the STS content of each clone defined the human chromosome 22q breakpoint position to within 1 kb between STS markers B1 and B2.

Because the majority of the fosmid clone fF45C1 had been sequenced at the time of this work, STSs could be defined and primers designed to assay for them. In order to test the primers for their specificity, hamster genomic DNA was included as a control because the human chromosome 22-hybrid DNA was isolated from a human x hamster hybrid cell line. If the primers generated product from human chromosome 22-hybrid DNA, but not from hamster genomic DNA, it was assumed that the primers were amplifying human chromosome 22 specific sequences. The fact that primers for A1, B16, B22 and B36 generated more than one product from human

genomic DNA was possibly due to the repetitive nature of the sequence in that region of the clone.

When assayed with Siamang DNA, 90% of the fF45C1 primer pairs generated PCR products. Those results implied that the particular region of Siamang chromosome 18 being investigated shares considerable sequence homology with human fosmid fF45C1. Not only is the homology close enough to allow the hybridisation by FISH of fF45C1, but also to allow the amplification of Siamang STSs using fF45C1 PCR primers.

Although 10% of the primer pairs failed to generate specific product from the gibbon, long range PCR was carried out to amplify overlapping sequences from Siamang DNA. The assumption that priming would be successful in the gibbon except when the primers lay either side of the breakpoint was a big one, especially as some of the primers failed under standard PCR conditions. Nonetheless from the results of the long-range PCR it was possible to exclude 24 kb of fF45C1 sequence as the location of the breakpoint.

Because the long-range PCR approach did not provide a definitive answer in locating the position of the breakpoint and limited the exclusion to only 24 kb of sequence, it was necessary to apply an alternative approach to identify and analyse the breakpoint. Therefore, a genomic gibbon library was constructed and screened for clones containing sequences spanning the fusion points.

After analysing the high molecular weight Siamang genomic DNA using *Mbo* I and *dam* Methylase, it seemed to behave in a similar way to human DNA, and a competitive restriction digestion was set up to cut the DNA prior to cloning. The Siamang DNA cut much more readily than had been anticipated. It was only possible to generate the correctly sized fragments by titrating the *Mbo* I in a limiting enzyme restriction digestion. After screening the low-density library

filters by hybridisation the colonies selected from the regions generating large positive signals were verified by STS-PCR

There are two possible explanations for the results observed following STS-PCR analysis of the gibbon cosmids. One explanation could be that each of the four cosmids had their ends lying in the 1 kb region between B1 and B2, although this seems unlikely. The other explanation could be that the sequences spanning the rearrangement junctions lie in the 1 kb region between B1 and B2, which could be why none of the four clones contain both STSs B1 with B2.

From the STS maps, it could be seen that clones SCHSY4.11, 7.21 and 10.11 contained at least 5.4 kb of DNA homologous to human chromosome 22q13.1. Clone SCHSY4.21 contained at least 31 kb of DNA homologous to human chromosome 22q13.1 proximal to the breakpoint. The fact that, after FISH, clone SCHSY4.11 hybridised to Siamang 18p, human chromosome 22q13.1 and 16p confirmed that it contained sequences spanning the homologous block fusion point. Similarly, SCHSY4.21 hybridised to Siamang 18q, human chromosome 22q13.1 and 16p, confirming that it also contained sequences spanning the homologous block fusion point.

Having identified the two cosmid clones containing sequences spanning the rearrangement junctions in Siamang chromosome 18, it was necessary to study them at the sequence level in order to analyse the underlying structure and mechanism, which might have lead to the evolutionary chromosome rearrangement. The work towards that analysis is described in Chapter 6.