4 Periodic Flexibility Patterns in DNA: a Scan for Signals Involved in Nucleosome Translational Positioning

4.1 Introduction

Some recent computational approaches have indicated periodic occurrences of flexibility patterns in the range of 100-200 bp in eukaryotes but not in prokaryotes (Section 1.11.2). This suggests that these flexibility patterns could be involved in positioning nucleosomes, owing to their size which is of the size order of a nucleosome core particle (146 bp). This made it interesting to examine where such flexibility patterns are located with respect to gene features in eukaryotic genomes. The availability of mouse genomic sequences, particularly syntenic regions shared with human, was a benefit to this investigation as it could also be investigated whether such potential translational positioning signals were a general mechanism conserved in evolution. The approach taken was to use the wavelet tool (Section 2.4.1) to analyse the occurrences and distribution of flexibility patterns in genomic sequences.

4.2 Methods

4.2.1 Construction of flexibility sequences

Flexibility sequences (Section 2.3.1) were used to represent DNA as sequences of conformational flexibility values.

4.2.2 Wavelet transform of whole chromosomal flexibility sequences

Wavelet transforms were performed on whole chromosomal flexibility sequences using the software *Autosignal* (Clecom, 1999). The *Morlet* family of wavelets was used. This wavelet family is considered 'crude' in the respect that once transformed, the original data cannot be reliably reconstructed. However, signal reconstruction was not required in this analysis. The *Morlet* was an appropriate family to use for transforming flexibility sequences as it is suited for decomposing continuous data series such as flexibility sequences. The particular implementation of the *Morlet* family that was used was also a fast one, which calculates the Fourier transform of both the *Morlet* waveform and the raw signal (flexibility sequences) to achieve fast convolution.

The main datasets that were transformed and analysed were[18]:

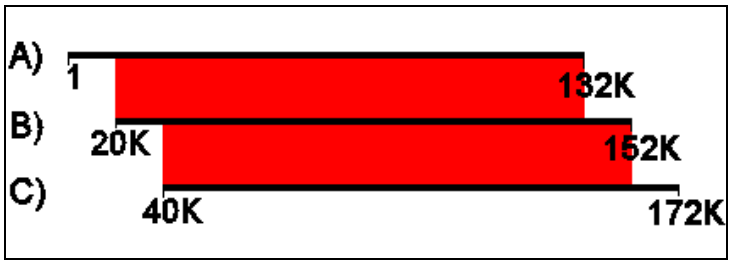
- Mycobacterium tuberculosis (Genbank ID: AE000516),
- Saccharomyces cerevisiae (Genbank ID: NC_001147),
- Human chromosome 20,
- Human chromosome 22,
- Mouse chromosome 19,

[18] Human and mouse data were extracted from the Ensembl database (Clamp *et al.*, 2003; Hubbard *et al.*, 2002)

- A 30MB syntenic region between human chromosome 20 (29.4MB to 62.9MB) and mouse chromosome 2 (172.1MB to 202.3MB).
- BRCA2 syntenic region between human and mouse (a 1.2 MB sequence alignment)

The period range which was analysed was 50-1000 bp; this range was selected such that periodic patterns of the length order of the nucleosome core particle (~146 bp) could be detected. Due to memory limitations as well as the software design constraints, the maximum sequence length that could be transformed at a time was 132,000 bp. Therefore, to handle chromosome-size data which covered several MB, a windowing scheme was used. Apart from the maximum data size, another limitation was the occurrence of edge effects associated with this wavelet family. These would result in a large amount of false classification towards the window edges. Therefore, an overlapping windowing scheme was adopted to minimize these effects (Figure 4.1). The start of each window was offset by a small amount (20,000 bp) relative to the size of the full analysis window (132,000 bp). So, for instance in Figure 4.1, strong patterns between co-ordinates 40,000 bp and 132,000 bp would only be considered if they appeared in all 3 analysis windows A, B and C.

Figure 4.1: Overlapping windowing scheme for removing edge effects in 'wavelet transform' analysis windows.



4.2.3 Thresholding by wavelet co-efficient strengths

The wavelet co-efficients, which represent the strength of a specific period along a flexibility sequence, are complex numbers. For purposes of visualisation and thresholding, these values were converted to decibels (dB) in *Autosignal*. This is measured as:

$$10.0 \times \log_{10}(r^2+i^2)$$

where *r* and *i* are the real and imaginary components of the wavelet co-efficients respectively. The strongest co-efficients, thus obtained in chromosomal flexibility sequences, were around 30.0 dB and the weakest were around -248.0 dB (0.0 dB is considered to be the lower limit for comparing 2 signals). 2D contour maps of the strengths of different wavelet co-efficients were plotted as in Figure 4.2 (page 4-122). For visualising the locations of strong patterns on sequences longer than the size of the wavelet analysis window, only regions stronger than 28.0 dB were plotted (for example, Figure 4.3, page 4-124).

4.2.4 Probability distribution of periodic flexibility patterns

The probability of observing a flexibility pattern, corresponding to a specific repeat period in the genome, was calculated as the total length occupied in a chromosome by such patterns divided by the total length of the chromosome. This was done separately for both introns and intergenic regions (for example, Figure 4.4, page 4-126).

4.2.5 Estimation of genomic features frequently associated with periodic flexibility patterns

The ratio of observed to expected frequencies was used to indicate which genomic features were frequently associated with flexibility patterns. The same procedure was used earlier (Section 3.2.9).

4.2.6 Alignment of flexibility sequences

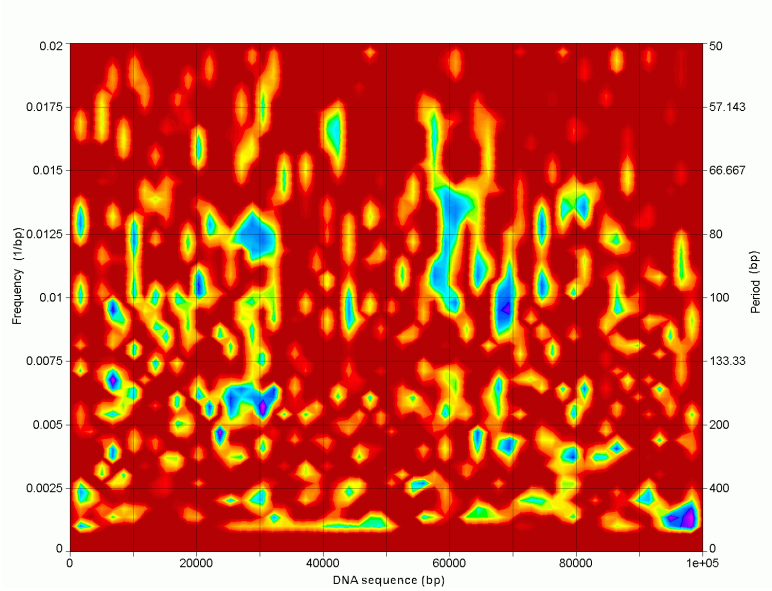
Sequences were aligned by their flexibility values in regions where strong wavelet co-efficient strengths (>28 dB) were obtained. A flexibility-sequence dataset was constructed by trimming 300 bp fragments around such regions. Following this, one sequence from this dataset was chosen randomly as a reference sequence. All other sequences were rotated until the offset of these sequences, having the strongest correlation co-efficient with the reference sequence, was found. The strongest offset flexibility sequences were then clustered and plotted as in Figure 4.6, page 4-129.

4.3 Results and Discussion

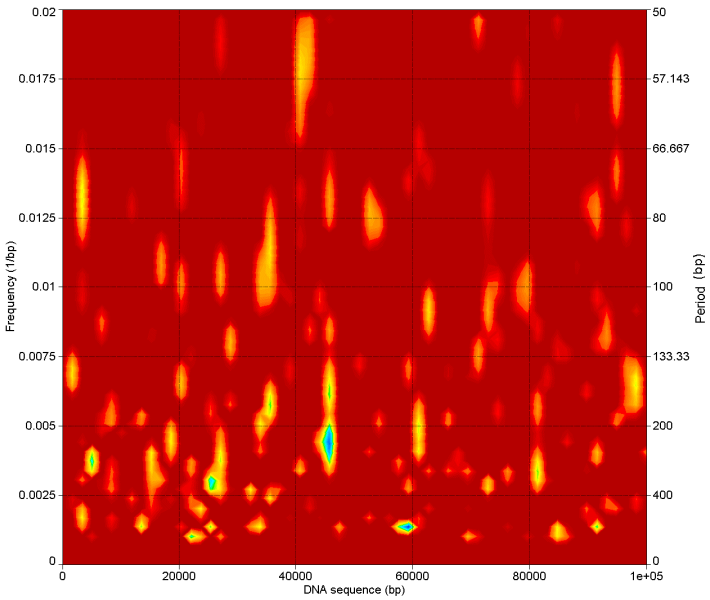
4.3.1 Differences in wavelet spectra between eukaryotic and prokaryotic flexibility sequences

The lack of nucleosome formation in prokaryotic genomes and their ubiquitous distribution in eukaryotic ones provides a reasonable basis for comparing their flexibility landscapes. To investigate this, 100 kbp-long flexibility sequences from human, *Saccharomyces cerevisiae*, and *Mycobacterium tuberculosis* were randomly selected and broken down using wavelet transformation (Figure 4.2). It was observed that in human, there was a dense distribution of periodic flexibility patterns, which was periodic between 50-1000 bp (Figure 4.2(a)). However, such patterns were not seen in *Saccharomyces cerevisiae* (Figure 4.2(b)) or in *Mycobacterium tuberculosis* (Figure 4.2(c)). Whereas the wavelet co-efficients in the human flexibility sequences were as high as 32 dB, the highest observed in *M. tuberculosis* or *Saccharomyces cerevisiae* was 24 dB. In the latter 2 genomes, there was still some weak periodicity, which was distributed sparsely. This distribution was not as dense as the stronger patterns seen in human. Upon completely randomizing the DNA sequence of the human clone and performing the wavelet transform on the corresponding flexibility sequence, the strong peaks were diminished yielding co-efficients which were now as high as 22 dB (data not shown). The lack of periodic flexibility patterns in *Saccharomyces cerevisiae* suggested that if such patterns did influence nucleosome positioning, then they would probably do so only in higher eukaryotic species.

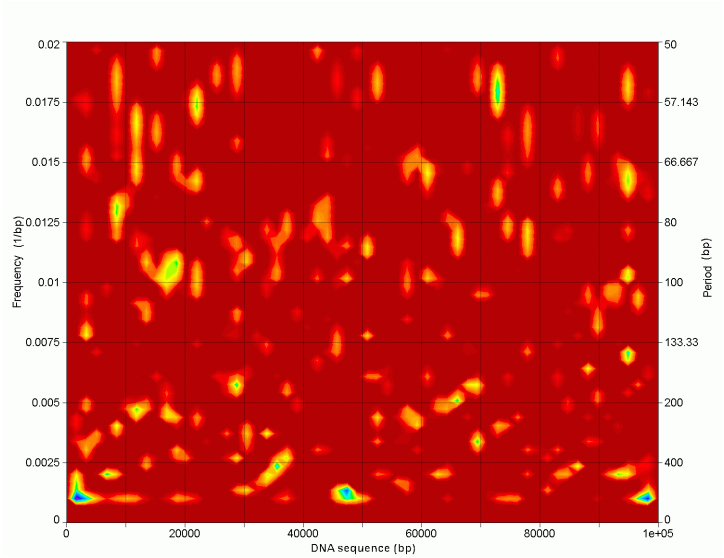
Figure 4.2: Continuous wavelet transform spectra compared between eukaryotic DNA flexibility sequences and a sample prokaryotic DNA flexibility sequence. The figures were obtained by performing the wavelet transform on randomly chosen 100,000 bp segments of the following sequences: (a) a clone from human chromosome 22 (Ensembl ID: AC004019.20.1.260409), (b) *Saccharomyces cerevisiae* chromosome XV (Genbank ID: NC_001147) and (c) the *Mycobacterium tuberculosis* genome (Genbank ID: AE000516). The units on the z-axis were measured in decibels (dB); the colour gradients shown are based on a contour map of 48 colours ascending from red to blue. Red represents 0 or <0 dB intensity and dark blue represents the strongest observed intensities around 31 dB.



(a)



(b)

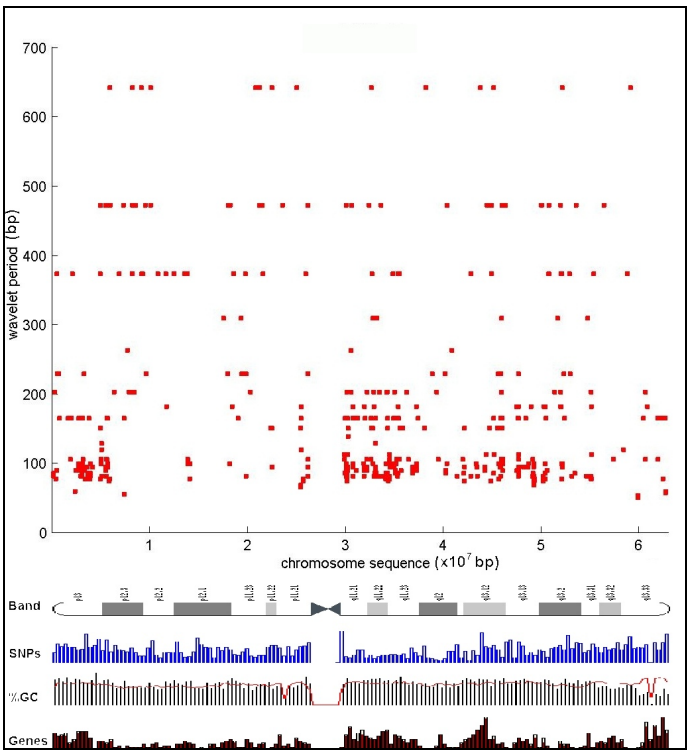


(c)

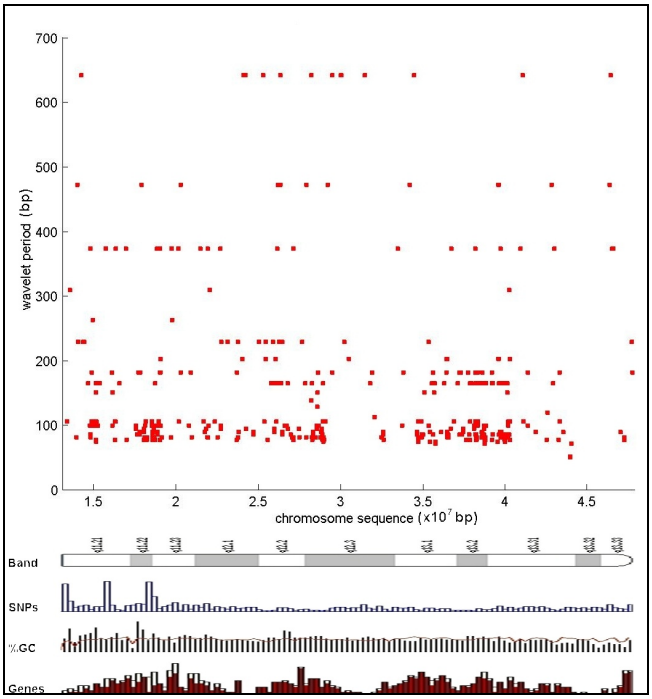
Such an examination of the flexibility landscapes of eukaryotic and prokaryotic DNA had been done before (Audit *et al.*, 2001; Audit *et al.*, 2002) utilising a different flexibility model (Goodsell & Dickerson, 1994). Using the wavelet technique to estimate a parameter called the Hurst exponent, Audit *et al* estimated that the occurrence of long range correlations of the order 10 – 200 bp was strong in several eukaryotic genomes including *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and human as well as in some of the viral genomes which infect them. The results obtained for *Saccharomyces cerevisiae* above, however, contradict this observation. They had also noted the lack of strongly periodic features in this range in bacterial genomes such as *Aquifex aeolicus* and *Bacillus subtilis*.

4.3.2 Whole chromosomal flexibility landscape in higher eukaryotic chromosomes

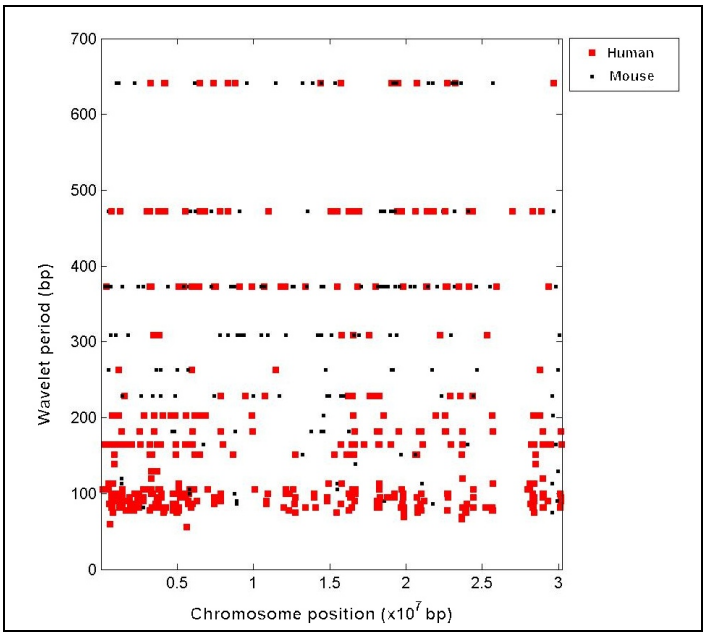
Figure 4.3: Continuous wavelet transforms of 3 large eukaryotic DNA contigs. These 2D plots were obtained by thresholding the wavelet co-efficients at 28 dB and plotting only those regions which were above this threshold. These results were obtained from transforming (a) 63 MB of human chromosome 20, (b) the q arm of chromosome 22 (32 MB) and (c) a 30 MB syntenic region between human chromosome 20 (sequence co-ordinates 29.4 MB to 62.9 MB) and mouse chromosome 2 (sequence co-ordinates 172.1 MB to 202.3 MB).



(a)



(b)



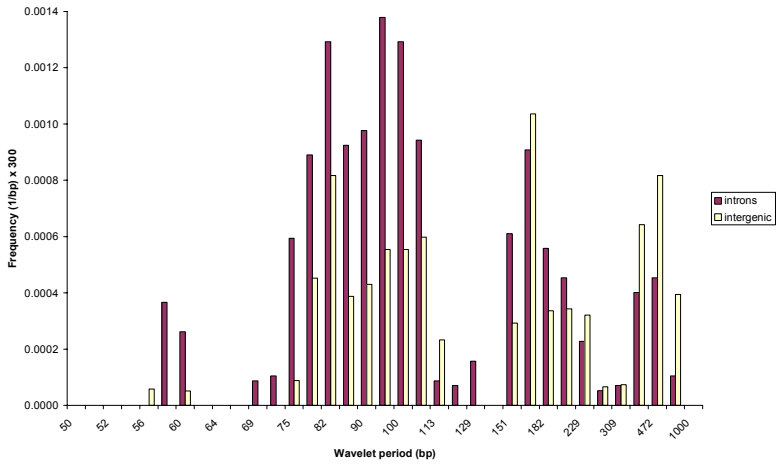
(c)

Figure 4.3(a),(b) shows the flexibility wavelet spectrum in relation to gene density in 2 human chromosomes. Distinct clumps of periodic flexibility patterns, in the range of 80–120 bp, were observed. In addition to these, there was a slightly less dense distribution of patterns observed in the range of 120–200 bp. The locations of these two "periodic classes" appeared to roughly coincide. Periodic patterns, above the 200 bp scale, occurred relatively sparsely.

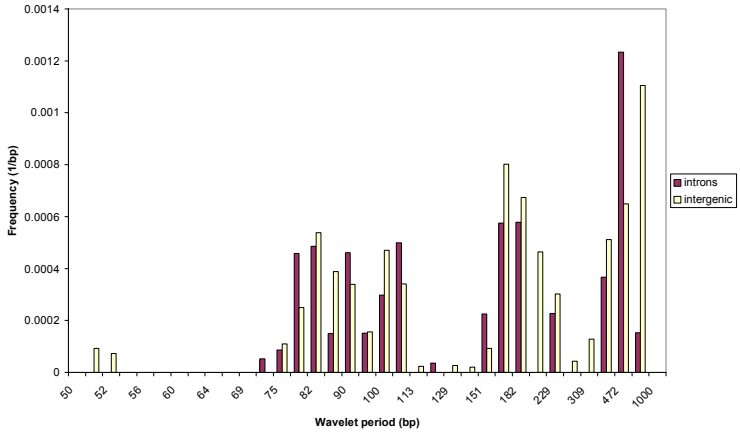
The dense clumps of 80–120 bp patterns also appeared to roughly coincide with gene density (Figure 4.3(a),(b)). This closeness was apparent along the following co-ordinates:

- Human chromosome 20 (Figure 4.3(a)): 1–7 MB; 30–38 MB; 40–50 MB
- Human chromosome 22 (Figure 4.3(b)): 17–20 MB; 25–30 MB; 35–40 MB

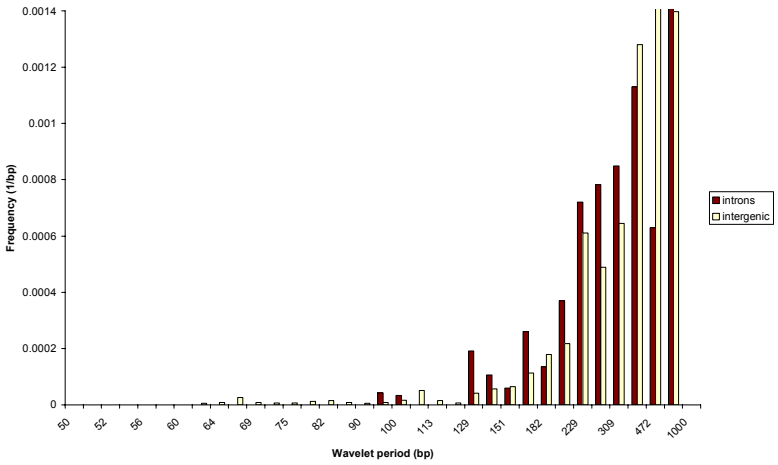
Figure 4.4: Probability distribution of observing periodic flexibility patterns in the range 50–1000 bp in 3 different eukaryotic chromosomes. The results were obtained from (a) human chromosome 20, (b) human chromosome 22 and (c) mouse chromosome 19.



(a)



(b)



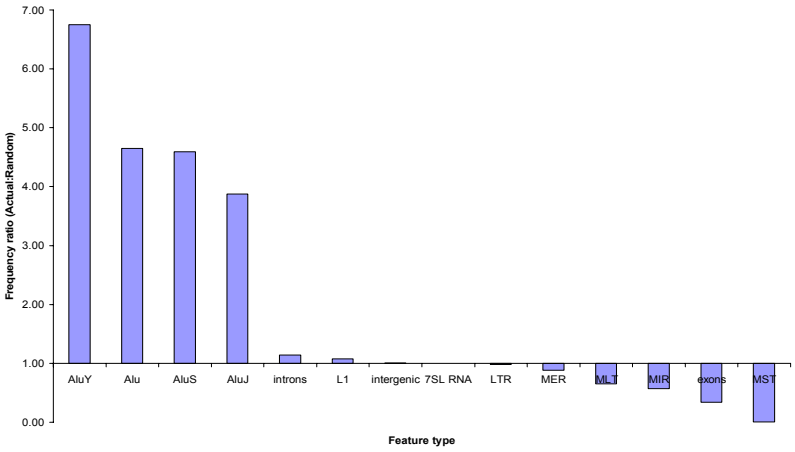
(c)

To gain further insight into the distribution of these flexibility patterns, the probabilities of observing each of the sampled periods were compared for the 2 human chromosomes (Figure 4.4(a),(b)). In both graphs, there were 3 distinct peaks visible, which corresponded to the 3 aforementioned "classes" of periodic patterns.

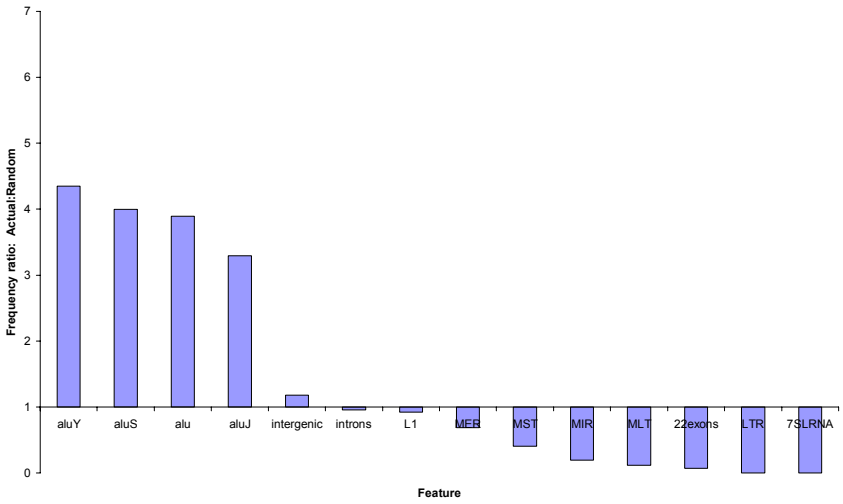
4.3.3 Genomic features frequently associated with strongly periodic flexibility patterns

The occurrence of strongly periodic flexibility regions could have simply been the result of recoding previously known eukaryotic DNA sequence elements. Especially given the fact that the periodic features aligned closely with gene-dense regions, this observation required a closer inspection.

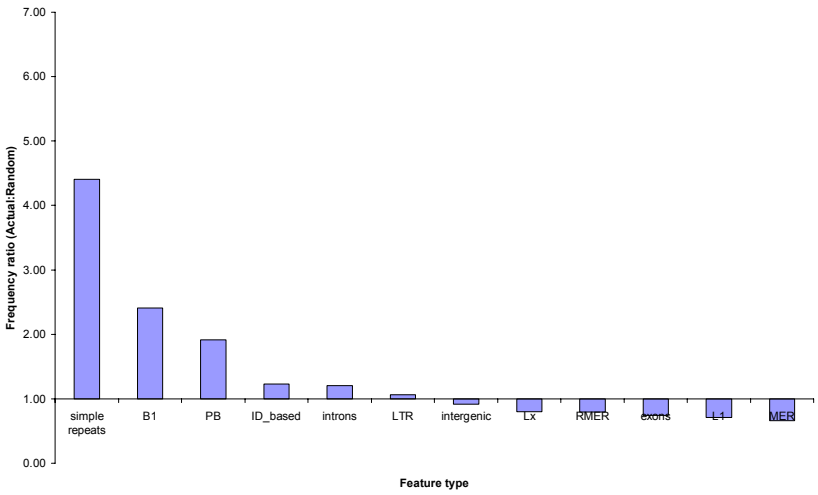
Figure 4.5: Features frequently associated with periodic flexibility patterns in (a) human chromosome 20, (b) human chromosome 22, and (c) mouse chromosome 19. Values greater than 1.0 indicate that a feature was more frequently associated with flexibility patterns than expected. The reverse is true for values less than 1.0.



(a)



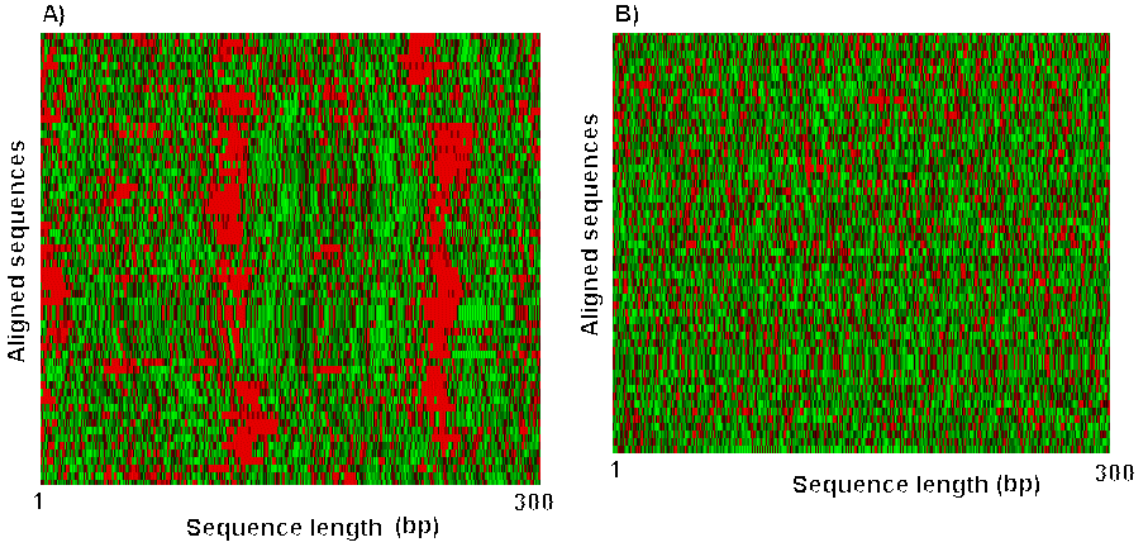
(b)



(c)

Figure 4.5(a,b) show features, which were frequently associated with the flexibility patterns of the 80–120 bp class in human. Clearly only the Alu repeat category was enriched: this repeat category was 4 times more frequently associated with these periodic flexibility patterns than expected. Aligning these sequences based on their flexibility (Figure 4.6) showed the linear arrangement of the periodic flexibility patterns that were detected. However, *RepeatMasker* analysis (Smit & Green, 1997) showed that the sequences themselves were mostly Alu repeats. So the observed patterns were in fact recoded Alu repeats (discussed in the next section; Section 4.3.4). Other notable observations from this analysis were that exons were not associated with these flexibility regions. This observation was consistent with other work, which suggests that long range correlations in eukaryotic DNA sequences exist only in non-coding DNA and not in coding sequences (Arneodo *et al.*, 1995; Arneodo *et al.*, 1998; Buldyrev *et al.*, 1998; Havlin *et al.*, 1999; Pattini L, 2001).

Figure 4.6: Flexibility alignment of 300 bp sequences of A) regions exhibiting 100–200 bp periodicity in flexibility (wavelet co-efficients >28 dB) and B) randomly selected human DNA sequences. Red and green colours represent strong rigidity and strong flexibility respectively. RepeatMasker analysis showed that the sequences in A) were mostly Alu repeats.



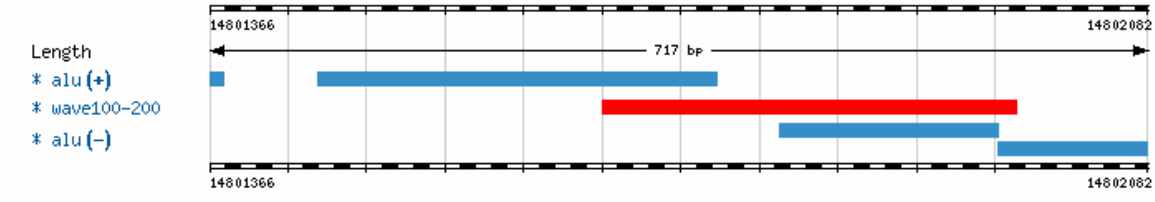
4.3.4 Why Alu repeats were frequently associated with periodic flexibility patterns

The results, discussed above (Section 4.3.3), showed that the flexibility patterns that were observed contained a large proportion of Alu repeats. The structure of Alu repeats themselves (Batzer *et al.*, 1996), as well as their recently outlined insertion patterns (HGSC, 2001), could explain why they had been detected using the wavelet transform.

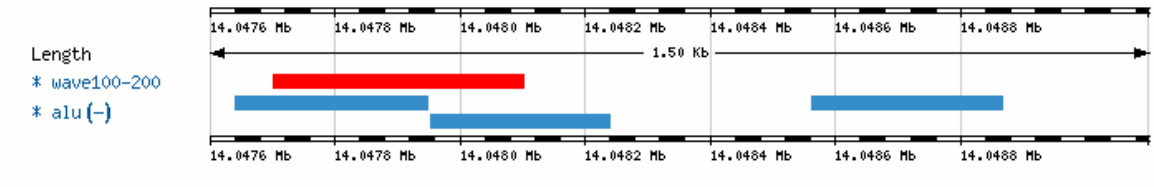
- **Alu structure**

Firstly, Alu repeats are dimers of two roughly 100 bp-long 7SL-RNA derived fragments (Batzer *et al.*, 1996); however, the left and right monomers do not share any sequence similarity. Alu sequences also contain a poly [A] linker region separating the 2 RNA fragments and a poly [A] tail at the 3' end. The tetranucleotide parameter set that was used for converting DNA sequences into flexibility (Section 2.3.1) and indeed most of the other DNA flexibility parameter sets (Bolshoy *et al.*, 1991; Brukner *et al.*, 1995; Goodsell & Dickerson, 1994; Olson *et al.*, 1998; Packer *et al.*, 2000a; Packer *et al.*, 2000b) all model poly [A] motifs as being rigid in conformation (Section 1.4.1). Therefore, in the flexibility sequences, which were supplied as input to the wavelet algorithm, the 100 bp–spaced poly [A] motifs were becoming recoded as 100 bp-spaced rigid motifs. However, the wavelet transform only yields strong co-efficients when there are locally periodic patterns. A more detailed view of such locally periodic regions (>28dB co-efficient strength) showed that Alu repeats, which were in a very close arrangement, accounted for the regions of high flexibility (Figure 4.7). This would explain the periodic patterns that were observed. The fact that Alu repeats could represent the major class of poly [A] sequences in human was indicated in much earlier work (Lustig & Petes, 1984).

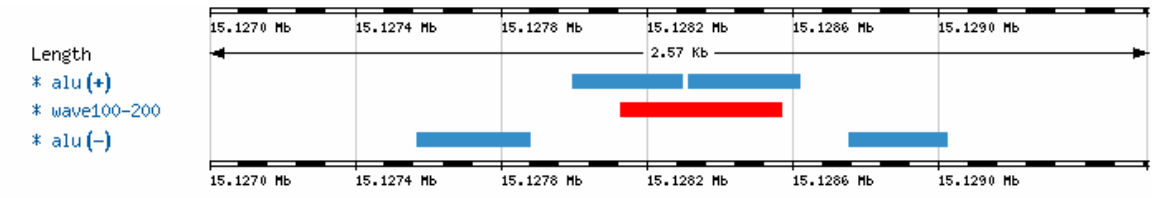
Figure 4.7: Zoomed view of periodic flexibility patterns (80–120 bp) having wavelet coefficient strengths >28 dB. 3 different resolutions are shown; in each case, the locally detected periodic flexibility is shown as a red bar. Positive and negative strand Alu repeats are shown as blue bars.



(a)



(b)



(c)

- **Alu retention biases**

Alu repeats have been reported to be preferentially retained in GC rich regions (HGSC, 2001). Although it is thought that Alu insertion is more or less random, it appears that they tend to remain fixed in GC rich regions (Smit, 1999). It had also been reported that most of the preferred GC rich regions were mostly occupied by the older[19] AluS. Younger Alu repeats were reported to be found in similar proportions in AT rich regions as GC rich regions possibly due to saturation of the GC sites by the older Alus (HGSC, 2001). Since genes also display a bias towards GC rich regions in the genome, it was apparent why the locations of strong periodic flexibility patterns and gene dense regions appeared correlated (Section 4.3.3). These results could also

[19] AluY are estimated to be 20 million years old; the middle aged Alus (aluS) 35 million years old; the oldest Alus (aluJ) 50 million years old (Batzer & Deininger, 2002)

explain Arneodo's observation that long range correlations in human DNA were related to GC content (Arneodo *et al.*, 1998).

- **Percentages of repeat families picked up by the wavelet transform**

To estimate whether the patterns picked up by the wavelet transformation were representative of the whole population of Alu sequences, the percentages of different repeat families associated with periodic flexibility were compared (Table 4.1(a)). As seen in Table 4.1(a), only 2.06 - 2.67% of any of the Alu age categories were detected as strongly periodic flexibility patterns. However, roughly 82.07% of the regions classified as highly periodic were associated with Alu sequences of all ages. Therefore, although the wavelet transform itself was strongly biased towards picking up Alu sequences, the total Alu population which they had picked up represented only a small fraction of the total Alu population (presumably only the ones whose positions were very close to each other). L1 repeats were also represented as highly periodic flexibility regions (25% in Table 4.1(a)); this could once again be due to the wavelet transform picking up clustered Alu repeats, which are thought to rely on the endonuclease activity of L1 repeats for their own replication.

Table 4.1: Percentages of repeat families which were associated with strongly periodic flexibility regions (wavelet co-efficients >28 dB) in descending order. These are compared to the proportion of total observed periodic flexibility (second column). The second columns do not sum to 100% as the proportion is measured across the distribution of a range of periodic patterns (for instance, the same region may be strong for both 80 bp periodic as well as 200 bp periodic patterns).

(a) Human chromosome 20

	% repeat	proportion of total periodic flexibility
aluY	2.67	18.35
aluJ	2.17	31.01
alu	2.16	82.07
aluS	2.06	62.66
LTR	1.36	3.59
MIR	1.99	5.49
L1	1.13	25.95
MST	1.02	1.05
7SL RNA	1.02	0.42
MER	0.85	9.49
MLT	0.51	3.59

(b) Mouse chromosome 19

	% repeat	proportion of total periodic flexibility
Simple sequence repeats	1.98	63.01
MER	1.40	1.83
RMER	1.38	2.74
B-type	1.38	44.00
PB1	1.21	9.36
Lx	0.92	10.50
L1	0.89	11.87
ID-based	0.81	5.94

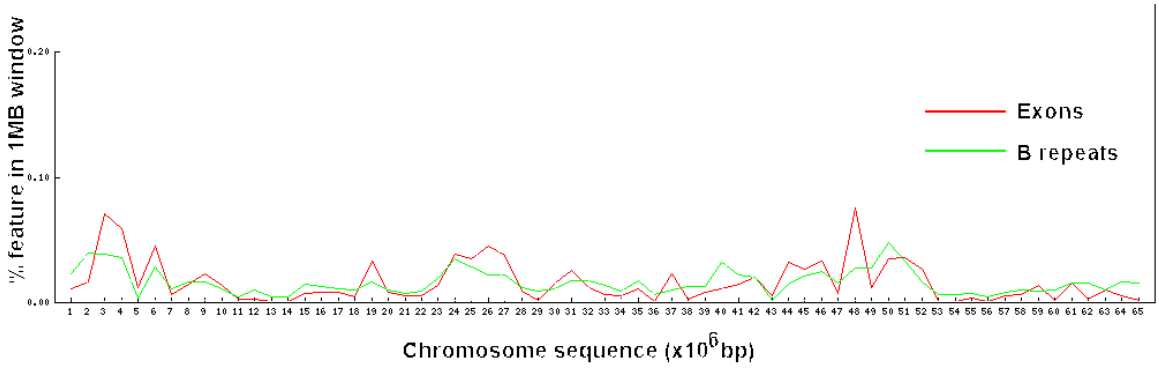
4.3.5 Conservation of periodic flexibility patterns in eukaryotic genomes

An important feature of a nucleosome positioning pattern may be that it is highly or at least moderately conserved between 2 species. To investigate this, a similar investigation, using wavelet transformation of flexibility sequences, was performed on mouse genomic contigs as was done for the human contig data. The data for the

mouse genome was available only during the latter stages of this analysis; the data was, therefore, in its infancy and not as refined as the analysed human contigs. The only high quality alignment between human and mouse available at the time was the BRCA2 syntenic region (a 1.2 MB sequence alignment). Similar flexibility patterns were not observed between human and mouse in the BRCA2 syntenic region though (data not shown).

Figure 4.3(c) (page 4-124) shows the results of applying the wavelet transform on flexibility sequences in a syntenic region in human and mouse. The densities of periodic patterns that were observed in mouse were much lower compared to those in human. The locations of such patterns also did not show any kind of similarity with any corresponding locations in human. Furthermore, the probabilities of observing the different periodic patterns were not similar to what was seen in human (Figure 4.4(c), page 4-126). The peak periodicities in human could be grouped into 3 distinct classes but in mouse, there was only a single broad peak with a maximum of around 600 bp. These results indicated that the periodic flexibility patterns, which were seen in human (and which largely resulted from the clustering of Alu repeats), were not conserved in mouse.

Figure 4.8: Correlation of B repeat density and gene density in a region of mouse chromosome 2.



Genomic features, frequently associated with these periodic flexibility patterns, were found to be mainly simple repeats and B1 repeats in mouse (Figure 4.5(c); Table 4.1(b)). Whereas in mouse, simple repeats accounted for roughly 63% of the total periodic patterns represented (Table 4.1(b)), in human, simple repeats were only marginally picked up by the wavelet transform: these peaked at 50 bp periodicity and there were 2 such regions near the telomeric regions of both human chromosome 20 and 22 (data not shown). B1 repeats are the lineage specific SINE family in mouse, which are monomers of roughly 100 bp and similar in sequence to the left monomer of Alu repeats (Quentin, 1994). They also show a bias towards being retained in GC rich regions (alongside gene dense regions) (Figure 4.8), a pattern which was pointed out in the recent analysis of the mouse genome (IMGSC, 2002). Therefore, B1 repeats, although they show the same biased retention patterns as their human counterpart, do not represent the same contribution of periodic rigidity in mouse. This result is expected from the inherent structure of B1 repeats, which are monomers and do not share the poly [A] linker and poly [A] tail motifs of their human counterparts. Similar to the lack of periodic flexibility observed in human exons, mouse exons also lacked periodic flexibility behaviour (Figure 4.5(c), page 4-128).

4.3.6 Re-examination of the hypothesis of nucleosome translational positioning with respect to Alu repeats

The current research has raised a fundamental question: "Is it likely that Alu sequences direct the translational positioning of nucleosomes?". Although a conclusive answer cannot be provided owing to the limits of the methodology outlined in this chapter, the evidence obtained using independent methods which link Alu repeats with nucleosome positioning can be considered. Secondly, there is also significant evidence in the literature that suggests that Alu sequences have acquired

important functional roles in the human genome. Although these functional roles may not be directly related to nucleosome positioning themselves, the critical nature of the functions themselves may influence opinion on whether Alu sequences have developed a close enough symbiotic relationship in the host genome that could include effects such as nucleosome positioning.

- **Other evidence linking Alu sequences and nucleosome positioning**

The only recent computational work, which had connected Alu repeats and nucleosome positioning, was using the measurement of dinucleotide relative abundance distance discussed earlier (Section 1.9.4). This concluded that Alu repeats had the highest nucleosome formation potential but the nucleosome model used was itself questionable.

Fox *et al* (Fox, 1992) reported that large-scale isolation of genomic poly [A] clones (containing a large amount of Alu sequences) and reconstitution onto nucleosome core particles did not show significant aversion to nucleosome binding compared to random DNA fragments. This result was contradicted later by Englander (Englander *et al.*, 1993), who showed that Alu sequences showed rotational and translational nucleosome positioning capacity using an *in vitro* nucleosome reconstitution experiment. They showed that transcription in the *in vitro* DNA construct was blocked by nucleosomes; these nucleosomes were thought to be translationally positioned over the Alu elements. DNase I digestion indicated that the poly [A] linker region and poly [A] tails of the Alu sequences were probably directing this positioning. Englander *et al* later estimated that the left monomer of Alu repeats probably also had rotational positioning capacity (using DNase I digestion and software analysis) (Englander & Howard, 1995).

Englander *et al*'s results, particularly in (Englander & Howard, 1995), have interesting implications for the observations made in this chapter. Firstly, they estimated a rotational component in only the left monomer of the Alu sequences; this sequence is a homolog of B1 repeats in mouse (Quentin, 1994). This could suggest that clustering of Alu repeats and B1 repeats in GC rich regions ensures a significant quantity of rotational positioning signals in the upstream regions of genes in human and mouse respectively. This feature would not have been picked up in the current wavelet approach since the software they had used, for measuring curvature, was based on scanning for curved DNA; the wavelet tool used here was used to detect periodic flexibility of the scale order of 50–1000 bp. However, according to the signal which was picked up by the wavelet transform, namely the poly [A] motifs of the Alu repeats, it was highly unlikely that translational positioning was a conserved mechanism between human and mouse. The conclusion from linking the wavelet results from to Englander *et al*'s work is, therefore, an interesting one: rotational nucleosome positioning could be conserved between mouse and human but translational positioning is unlikely.

- **Alu repeats have taken on important functional roles in the cell**

One theory suggests that "Alu elements integrate randomly but those that are actively transcribed (and are therefore more likely to reside in G+C rich regions of the genome) are more likely to become fixed in the population " (Smit, 1999). This suggests that Alu repeats may play some functional roles due to their retention near gene dense regions (G+C regions). And indeed a number of recent experiments have shown that Alu sequences have adopted roles in important cellular functions.

Firstly, $1/3^{rd}$ of CpG islands have been estimated to be contained within Alu repeats (Rubin *et al.*, 1994; Schmid, 1991). This could suggest that Alu repeats have

an effect on the expression pattern of downstream genes due to mutations that alter the CpG methylation patterns. Alus are also known to directly insert into coding sequences and 0.1% of all genetic disorders are known to be caused by such unfavourable insertions (Deininger & Batzer, 1999).

In many organisms, SINE expression levels also increase under conditions of stress (Chu *et al.*, 1998; Li *et al.*, 1999; Liu *et al.*, 1995). Under such conditions, SINE RNA transcript has been reported to bind a specific protein inhibitor, and thereby block its activity. Therefore, under conditions of stress, Alu repeats may be specifically induced to upregulate the expression of many genes. This increase in Alu expression has also been linked with a rise in DNase I hypersensitivity in chromatin indicating possible Alu-mediated reshuffling of chromatin arrangement (Kim *et al.*, 2001).

Some recent work has provided the first indications of common functional roles between Alu and B1 repeats in human and mouse respectively (Zhou *et al.*, 2000; Zhou *et al.*, 2002). Zhou *et al* showed that the strongly evolutionarily conserved *Pax6* transcription factor, which is critical in the development of the eye, pancreas and central nervous system, exhibits more than 1 kind of preferred binding site in both human and mouse. However, the transcription factor binding sites included several Alu repeats in human and B1 repeats in mouse. An interesting twist was that the binding sites in the 2 lineage-specific SINE families did not share any sequence similarity! This suggests that the evolution of *PAX6* function may have been aided or merely influenced by simultaneous SINE evolution.

4.4 Conclusion

The wavelet transformation of flexibility sequences showed that the clustering of Alu repeats resulted in locally periodic rigidity patterns. On account of such clustering, two classes of periodicity could be seen: 80–120 bp and 120-200 bp respectively. These were observed near gene dense regions, which was expected from the biased retention property of Alu repeats in GC rich regions. Similar flexibility patterns were not seen in analysis of mouse contigs. SINE repeats may have simultaneously evolved to serve some common functions in their respective host genomes. But according to the results presented in this chapter, it is unlikely that nucleosome translational positioning is one such conserved function.