7. Discussion

In my thesis I have outlined strategies that allow for *in vitro* selection of torsionally strained DNA molecules. The validity of the strategies was proven for the case of Z-DNA:protein interactions. A Z-DNA binding protein was useful in this study, since under physiological conditions negative supercoiling is necessary for formation of Z-DNA. The Zα domain of the ADAR1 protein was chosen since Previous experiments indicated that this protein domain binds to DNA present in the Z-DNA conformation.

At the same time it was hoped to gain further insight into the manner in which the Zα domain binds to Z-DNA; especially if a preferentially bound sequence exists

7.1. Circular dichroism spectroscopy

First it had to be established, that Zα peptide actually does bind to Z-DNA and not to the B-DNA conformer. It has been shown previously that the Zα domain binds to a $(CG)_{35}\cdot(CG)_{35}$ oligomer stabilized in the Z-DNA conformation by bromination. Complex formation could be competed by negatively supercoiled plasmid DNA, containing a $(CG)_{12}$ sequence likely to adopt the Z-DNA conformation (Herbert et al. 1995). Here, further evidence is presented for the binding of the Zα domain to Z-DNA in the form of circular dichroism spectra and bandshift assays using unmodified DNA molecules.

Circular dichroism probes the conformation of chiral molecules. Circular dichroism spectroscopy has long been used for the investigation of DNA structure, and in fact the first observation of Z-DNA was made during a circular dichroism study. Pohl and Jovin observed an inversion of the spectrum of a solution of poly(GC)·poly(GC) under high salt conditions (Pohl and Jovin 1972). Later on, circular dichroism spectroscopy was used to investigate the effect of different salts on the B-Z equilibrium (Behe and Felsenfeld 1981). While the circular dichroism spectrum between 240 nm and 320 nm wavelength is a valid probe for DNA conformation, it does not constitute a final proof. A point in case is the compound Mitomycin C which, based on the circular dichroism spectrum, was proposed to cause DNA to adopt the Z-DNA conformation (Mercado and Thomasz 1977) (Wu et al. 1981). Further studies, however, showed that this was not the case. Treatment of DNA with Mitomycin C did not facilitate the conversion of B-DNA to Z-DNA in the presence of ethanol, nor did the treated DNA bind to antibodies directed against Z-DNA (Tomasz et al. 1983). The interaction of Mitomycin C with DNA was not dependent on the Z-DNA

conformation (Chawla and Tomasz 1988) and a high resolution structure by NMR of the Mitomycin C-DNA complex revealed a B-DNA conformation (Sastry et al. 1995).

The spectra obtained after addition of the Zα peptide to poly(m^5CG)·poly(m^5CG) showed the typical features observed in Z-DNA spectra : Loss of the negative band at 253 nm and appearance of a new negative band at 293 nm (Pohl and Jovin 1971). The isodichroic point observed during titration with Zα peptide (278 nm) is not the same as the one observed using $MgCl_2$ (272 nm). This is difficult to explain. However, the spectra published in the literature also show different isodichroic points ranging from 271 nm (Takeuchi et al. 1991) to 279 nm (Ulanov et al. 1992). The precise position of the isodichroic point may be dependent on small structural differences. Even though the circular dichroism spectrum at the wavelengths of interest is dominated by the aromatic bases of the DNA, it seems plausible that proteins or salt ions, interacting with the phosphate backbone or the bases, will cause differences in the circular dichroism spectrum.

7.2. Bandshift assays

The circular dichroism spectroscopic study was performed on a methylated polymer: In order to show that Zα peptide can bind to sequences not stabilized by chemical modification, bandshift assays with unmodified oligonucleotides were performed. Although bandshift assays can only provide indirect evidence for conformation specific binding, their high sensitivity is useful. In the past, bandshift assays with Z-DNA binding proteins have used poly(CG)·poly(CG) sequences stabilized in the Z-DNA conformation either by bromination (Herbert and Rich 1993) or methylation (Zhang et al 1992). Due to the high affinity of Zα peptide to Z-DNA I decided to extend these studies to oligonucleotides that were not stabilized in the Z-DNA conformation by chemical modification. Under these conditions, the vast majority of unmodified oligonucleotides will be in the B-DNA conformation. However, the small fraction of molecules present in the Z-DNA conformation is sufficient to give rise to a shifted band.

One problem encountered in pursuing these experiments was the occurrence of a broad smear between the shifted band and free DNA. This was caused by protein:DNA complexes dissociating during the electrophoresis. They are unlikely to reform since the DNA molecule quickly reverts back to the B-DNA conformation. Therefore the duration of the electrophoresis had to be kept as short as possible. Hence, a comparatively high

voltage of 500 V was applied. The high voltage, in turn, required an electrophoresis apparatus in which the gel was immersed in buffer and thereby cooled.

A number of different DNA oligonucleotides were used as DNA substrates in bandshift assays (Figure 6.4). Most of these were hairpin molecules in which two reverse complementary sequences were connected to each other by a loop of four deoxythymidine nucleotides. Hairpin molecules form predominantly monomers in which the complementary bases are base paired. Due to an entropic effect this conformation is very stable, and the melting point is considerably higher than that of a double helix made up of DNA strands not covalently connected to each other (Data not shown). The $(T)_4$ loop does not interfere with the B-Z transition of the double-stranded stem and the molecule $(CG)_3T_4(CG)_3$ has been crystallized with the stem being present in the Z-DNA conformation (Chattopadhyaya et al. 1988).

Hairpin molecules with double stranded regions as short as a $(CG)_3{\cdot}(CG)_3$ motif were bound by the Zα peptide (Figure 6.4). Likewise binding was observed when a core sequence of three alternating GC or CG dinucleotides was flanked by overhangs of one or two unpaired nucleotides. Inhibitory to binding were interruptions of the alternation of purine and pyrimidine bases as was the case in the sequence $(GC)_3(CG)_3T_4(CG)_3(GC)_3$. Why the sequence $(GC)_3ACGCT_4GCGT(GC)_3$ did not give rise to a bandshift remains a mystery, especially since the sequence $AC(GC)_3T_4(GC)_3GT$ did. The sequences $(GT)_6T_4(AC)_6$ and $(AT)_6T_4(AT)_6$ did not give rise to shifts. Both sequences do not contain any CG or GC dinucleotide sequence motifs. While AT dinucleotide motifs form Z-DNA poorly (Ellison et al. 1986), GT/AC dinucleotide motifs are compatible with the formation of Z-DNA. Similar results have recently been published, there a $(CA)_3T_3(TG)_3$ hairpin did not give rise to a bandshift with Zα peptide (Herbert et al. 1998). Only after extending the sequence by a $(CG)_3$ motif was a bandshift observed. Binding of Zα peptide to the $(CG)_3$ motif was apparently sufficient to stabilize the whole sequence in the Z-DNA conformation. Aa second molecule of Zα peptide was then able to bind to the $(CA)_3$ motif. The same was observed for a $(TA)_3(CG)_3T_3(CG)_3(TA)_3$ hairpin molecule. It appears that the inability to obtain a bandshift with a pure GT/AC sequence does not indicate that Zα peptide cannot bind to these sequences. Rather, such sequences are difficult to stabilize in the Z-DNA conformation.

In order to investigate the interaction of polynucleotides with Zα peptide, the complex of radioactively labeled $(CG)_4T_4(CG)_4$ hairpin and Zα peptide was competed with several

other nucleic acids. As expected, competition with the plasmids pME.E and pME.C, containing a $(CG)_{11}$ and $(TG)_{30}$ insert respectively, resulted in loss of binding to the hairpin probe. Competition with the plasmid pME.A, containing no such insert of alternating purines and pyrimidines, had only a limited effect (Figure 6.6). Whereas hairpins containing repeats of TG dinucleotide motifs did not give rise to a bandshift, the result obtained with plasmid pME.C indicates that Zα peptide can indeed bind to repeats of TG dinucleotides if present in the Z-DNA conformation. The plasmid pME.A is not expected to adopt the Z-DNA conformation under these conditions. Therefore, it is interesting to note, that, based on molar amounts, a 70 fold excess of plasmid over Z-DNA probe is sufficient for partial competition. However, if the competition reflects unspecific interaction of the Zα peptide with B-DNA, rather than short stretches of Z-DNA, then the relative concentrations are better expressed in mass units. Hence, a 1,300 fold excess of B-DNA did not compete binding of Zα peptide to a Z-DNA probe and a 13,000 fold excess results only in a partial competition. As was expected, binding of Zα peptide to the hairpin was competed by methylated and un-methylated poly(CG)·poly(CG) the two DNA sequences that most easily adopt the Z-DNA conformation.

Competition with different RNAs showed no effect unless very high concentrations of RNA were used. Even at high concentrations of RNA, only tRNA but not poly adenylic acid diminished the shifted band. The question whether the Zα domain can bind to RNA is of interest, since the only known biological activity of the ADAR1 protein is the editing of double-stranded RNA sequences. Furthermore, the Zα domain has some similarity to domains present in the E3L protein (Patterson and Samuel 1995) and the 2´-5´Oligoadenylate Synthetases (unpublished), both of which are proteins capable of binding to double stranded RNA. It has been observed that recombinant Zα, when purified from *E. coli*, copurifies with a small RNA whose sequence corresponds to a small stretch of the 16 S ribosomal RNA (T. Schwartz; personal communication). However, competition of a preparation of chicken ADAR1 with $r(CG)_{12}$ did not diminish the binding to a Z-DNA probe (Herbert et al. 1995). Likewise, longer dsRNA polymers such as poly[r(A·U)] or poly[r(I·C)] did not affect the binding to the Z-DNA probe. The latter RNAs led to a supershift which can be explained by their binding to the dsRNA binding motifs of ADAR1.

One of the goals of my thesis was to investigate whether the binding of the Zα peptide to DNA is influenced by factors other than the propensity of a sequence to form Z-DNA. A variety of different oligonucleotides were therefore labeled radioactively and used in

bandshift assays (Figure 6.8). Sequences of alternating purines and pyrimidines were bound by GST-Zα even when the stems of the hairpin molecules were just 6 nucleotides long. Binding of such a short sequences is consistent with a numerical analysis based on the circular dichroism data where the maximal effect was observed at a ratio of protein to DNA close to one molecule of Zα peptide per 5 basepairs of DNA.

7.3. *In vitro* selection

7.3.1. *In vitro* selection of plasmids

Determination of the optimal binding sequence by bandshift assay requires the synthesis of each individual sequence to be analyzed. This sets a practicle limit to the number of sequences that can be analyzed and hence assumptions have to be made as to what could be a good binding sequence. A better way to accomplish this task is *in vitro* selection, of which the principles, advantages and limitations have already been described in the introduction.

As mentioned, several prerequisites exist for the formation of Z-DNA. First of all, there has to be a stretch of alternating purine and pyrimidine bases. However, even in the presence of a suitable sequence the DNA will predominantly be present in the B-DNA conformation. Only stabilization of the Z-DNA conformation by high salt, chemical modification or torsional strain, leads to a noticeable shift in the B-Z equilibrium. Since it was intended to investigate the interaction of a protein with Z-DNA high salt conditions were not likely to be useful since the interaction of proteins with DNA is dominated by the contribution of electrostatic forces. These forces, however, would be weakened by conditions of high salt, resulting in the loss of binding. Chemical modification of the DNA would have been possible, however, a number of drawbacks exist. First of all, chemical modification does not reflect the *in vivo* situation. While methylation of cytosine bases is a frequent feature, especially in CG dinucleotide motifs, its significance for formation of Z-DNA *in vivo* is not known. Since only certain types of bases can be modified, a modification strategy might introduce a bias towards sequences containing these bases. Furthermore, the position and extend of modification would vary between selection cycles. Finally, no easy experimental procedures exist to precisely determine which bases are modified in the selected sequences.

Stabilization of the Z-DNA conformation by torsional strain therefore is the only feasible strategy. While the degree of supercoiling varies between species, the level of negative supercoiling present in plasmids isolated from the usual

grow on the agar plate. Using plates of 20 cm x 20 cm up to app. 100,000 distinct colonies per plate can be grown in such a way. A protocol has been published in which millions of clones were generated (Smith 1992). However, this necessitated the use of 64 large agar plates. It is obvious that with increasing numbers of individual clones the necessary experimental effort becomes impractical. By shortening the randomized region the necessary number of clones needed to fully represent al sequences can by reduced. However, a minimal length of alternating purines and pyrimidines is required to enable DNA to adopt the Z-DNA conformation. Sequence motifs of just 5 nucleotides length were bound in bandshift assays. However, the bandshift assays were done with hairpin oligonucleotides in which the whole double stranded part of the DNA molecule can adopt the Z-DNA conformation without the formation of B-Z junctions. The formation of the junctions is the energetically most costly step in the formation of Z-DNA. Therefore a stretch of Z-DNA in a plasmid has to be of sufficient length to pay for the energetic cost of two B-Z junctions. In view of both quantitative problems it was decided to use a randomized sequence of alternating purines and pyrimidines instead of a fully randomized sequence. This radically lowered the complexity of the library, meaning that less clones were necessary in order to fully represent all possible sequences. An initial attempt was made using a plasmid library containing a $(YR)_5$ sequence. However, selection experiments using this library were unsuccessful. It was concluded that the $(YR)_5$ insert was too short to permit the formation of Z-DNA.

A second library was constructed, using a $(YR)_8$ randomized sequence. Analysis of the library sequences before selection revealed that the randomized region often had undergone deletion of 2 nucleotides or multiples thereof. The observation that sequences with a high propensity to form Z-DNA are prone to deletions had been reported before. When $(CG)_n$ repeats of varying lengths were inserted into the plasmid pBR322 segments longer than 50 bp were not stable, whereas segments shorter than 30 bp were stable in most cases (Klysik et al. 1982). The frequency of deletions seemed to depend not only on the length of the insert, but also on the insertion site within the plasmid. In a different paper it has been reported that the insertion of Z-DNA forming sequences into the pUC8 plasmid also results in hotspots of spontaneous deletions, all of which involve an even number of base pairs (Freund et al. 1989). Sequences were prone to deletions until a size distribution centered around $(GC)_6$ was reached. The deletion events were classified as either small addition/deletion events (SADEs) which are additions or deletions of one or two dinucleotides, and long deletion events (LDEs) which are deletions of more than two

dinucleotides. In the case of pUC-$(CG)_{13}$ SADEs accounted for 10 % of the events and LDEs for about 90 % of mutation events. The SADEs were proposed to stem from slipped-strand mispairing of the two DNA strands within repeated sequences, with the B-Z junction triggering the deletion (Levinson and Gutman 1987). The deletion process did not appear to be dependent on the SOS DNA repair system, and was not affected by a *recA* mutation (Freund 1990). The observed mutation frequencies varied between 0.3 x 10^{-4} (pUC-$(GC)_6$) and 0.3 (pUC-$(GC)_{13}$). The deletions observed in the plasmid library probably are comparable to the SAD events. However, it is interesting to note the high frequency of deletion events observed in the plasmid library. The frequency of deletion is thought to depend on the location of insertion into the plasmid, which may explain the observed differences. However the sequences in this work were inserted into the same polylinker in which $(TG)_{30}$ and $(CG)_{11}$ inserts seem to be quite stable. The exact reason for the high frequency of the deletion events (0.5) remains unexplained.

Four rounds of *in vitro* selection using GST-Zα peptide led to a collection of sequences with interesting properties. First of all, almost all sequences contained at least part of a second block of (YR) dinucleotides. Whereas dinucleotide deletion events in the first block were not frequent, most of the second blocks were missing two or three dinucleotides. The two blocks were usually separated by a spacer whose sequence was variable, but related to a consensus sequence. The consensus sequence is a combination of the five nucleotides preceding the 5´ end of the randomized block, and the five nucleotides following the 3´ end of the randomized block. Conceivably, the duplication of the randomized block could have occurred either during the construction of the library, i.e. the PCR and the ligation reaction, or as a result of recombination once the plasmids were introduced into bacteria. Both the PCR and the ligation reactions seem unlikely sources for duplications since the linker and the second block of randomized nucleotides are highly variable between the sequences. It has to be assumed that some kind of recombination event took place, once the plasmids had been inserted into the bacteria. Interestingly, these recombination events had to work in a directed fashion, since almost always the second block shows dinucleotide deletions. However, the lack of similarity between sequences of the first and the second block argues against recombination. The second block therefore may not have been part of the same plasmid, but originates from recombination between two plasmids. The duplications described here have not been observed before. Furthermore, due to the use of strictly alternating sequences, previous studies have not been able to address the question of directionality in the deletion

process. Further investigation into the effects described here might help to identify the recombination process and the enzymes involved in the instability of Z-DNA in bacterial plasmids.

Analysis of the selected sequences revealed that the selected library is enriched in GC dinucleotide motifs and depleted of TA dinucleotide motifs, as would be expected for Z-DNA forming sequences. Likewise, analysis of longer sequence words also showed that sequences containing alternating CG dinucleotides are most common and those containing alternating TA dinucleotides least common. The longest sequence word present in all selected sequences is CGCGC/GCGCG . The only sequence words of four nucleotides length contained in all selected sequences are: ACGC/GCGT; CACG/CGTG; CGCA/TGCG; CGCG and GCGC. Of the possible sequences of three nucleotides length only half are represented in all selected sequences : ACG/CGT; CAC/GTG; CGC/GCG and GCA/TGC.

Does the plasmid selection give any clear evidence for the existence of preferred binding sequence of the Zα peptide ? The longest sequence word common to all selected sequences is the 5mer CGCGC . This is consistent with the results of the bandshift assays using mutated hairpin oligonucleotides. There, binding was significantly weaker or lost completely when this sequence motif was destroyed. However, at the same time CGCGC is the sequence with the highest propensity to adopt the Z-DNA conformation. This might be an indication for conformation specificity rather than sequence specific binding. Even more important : The statistical likelihood of finding the sequence CGCGC by chance in any particular one of the sequences in the library equals 80 % . Finding the sequence CGCGC in all the selected sequences is of fairly low statistical significance. As explained in the Appendix, sequence words shorter than 7 or 8 nucleotides cannot be detected due to the statistics of the $(YR)_8$ library.

Besides the observation that sequences rich in CG dinucleotides were selected, another observation argues that the main selection criterion was the propensity of the sequence to adopt the Z-DNA conformation. Almost all selected sequences contain a second block of randomized nucleotides. The second block was either not present at all in the original library or at low frequency at most. Therefore a strong selection pressure must have acted on these sequences for them to become dominant. Interestingly, at least half of the sequences contain two blocks of alternating pyrimidine/purine sequence interspersed by a spacer of up to eight nucleotides of non-alternating sequence. In principle two

interpretations are possible. The spacer sequence of non-alternating bases could have been forced into the Z-DNA conformation by the surrounding blocks of alternating purine and pyrimidine bases. Spacers of four bases length, enclosed by blocks of alternating CG, have been shown to adopt the Z-DNA conformation (McLean at el. 1986; Votavova et al. 1994). Whether this observation can be extended to spacers of eight bases length remains questionable, since very high salt was needed to flip these sequences to the Z-DNA conformation. Alternatively, the spacer sequence may remain in the B-DNA conformation, and closely spaced blocks of Z-DNA stabilize each other.

During the *in vitro* selection experiments using the plasmid libraries it became evident that contamination of the libraries by other plasmids can cause major problems. Several times the majority or even all the selected sequences turned out not to be plasmids from the library, but rather the plasmids pME.C and pME.E. These plasmids had been used as controls. Due to the superior ability of these plasmids to form Z-DNA, very small contaminating amounts of these plasmids were sufficient for them to dominate the entire library after a few selection cycles. This is further evidence that under the chosen conditions the propensity of a given sequence to form Z-DNA was the governing selection criterion. In order to overcome the contamination problem, the cloning strategy for the plasmid library was changed. The Sal I restriction site of the polylinker was destroyed when the randomized nucleotides were inserted. Since the contaminating plasmids pME.A and pME.E had been constructed leaving the Sal I site cleavable, restriction with endonuclease Sal I effectively eliminated the contaminating plasmids.

7.3.2. *In vitro* selection of minicircles

As mentioned above, the results of the *in vitro* selections using a $(YR)_8$ plasmid library did not yield a clear answer whether Zα peptide exhibits a sequence preference. Due to the statistical constraints of the plasmid library, only sequence words of more than 6 nucleotides would be significant. The results of the plasmid selection as well as circular dichroism spectroscopy and bandshift assays indicate that the Zα peptide interacts with about 5 consecutive nucleotides. The statistics of the plasmid library might have been improved by using a shorter randomized region. However a shorter region would have lowered the propensity of the sequence to adopt the Z-DNA conformation. Either the whole selection scheme would fail, or just the best Z-DNA forming sequences would be selected.

Since the necessity to transform bacteria had limited the achievable complexity of the plasmid library a strategy was developed that avoids the use of bacteria, and utilizes DNA minicircles instead. DNA minicircles are a powerful tool to study DNA structure, topology, and protein:DNA interactions. Minicircles allow for the stabilization of Z-DNA by negative supercoiling while, due to their smaller size, permitting the analysis of single Z-DNA tracts (Gruskin and Rich 1993). DNA minicircles have been used to investigate the interactions of Z-DNA with anti Z-DNA antibodies (Nordheim and Meese 1987) and Drosophila Topoisomerase II (Glikin et al. 1991). In the literature small amounts of radioactively labeled DNA minicircles have been used. Upscaling of the production protocols yields enough circular DNA to allow for *in vitro* selection experiments. A library of molecules containing 16 randomized nucleotides can contain up of 4^{16} = 4.3 x 10^{9} different sequences. This is equal to 7.1 x 10^{-15} mol . For a 580 bp minicircle this is equivalent to 2.7 x 10^{-9} g. In other words, a few nanogram of minicircle DNA would result in a fair representation of a N_{16} library. The minicircle size of 580 bp was chosen since a linear DNA of this length is most likely to be circularized in a ligation reaction (Shore and Baldwin 1983). Shorter sequences ligate less efficient, as the inherent stiffness of the DNA rod prevents the ends from meeting. Longer DNA molecules do also ligate less efficient since the likelihood of both ends meeting decreases with the length of the molecule. The minicircles used in this work are markedly longer than those used in previous work on Z-DNA molecules (260 bp).

Initially, a single PCR reaction was used to amplify the library between selection cycles. However, it was observed that, while the first selection cycle indicated specific binding of DNA to Zα peptide, subsequent cycles showed no such results. Only after the amplification protocol was changed to one similar to the one used for the generation of the initial library, was specific binding to the Zα peptide observed in later rounds of selection. Presumably, inefficient PCR conditions resulted in the DNA double strands being melted completely during the PCR cycle, without enough primer and/or polymerase being present to convert all melted strands into perfectly matched double stranded DNA. Upon cooling, the single stranded DNA molecules hybridized, however the randomized regions obviously remained mismatched. Another change introduced into the protocol was the use of a randomized sequence which contained a bias in favor of cytosine and guanosine nucleotides over thymidine and adenosine nucleotides. Since alternating CG or GC dinucleotides are known to stabilize the Z-DNA conformation, it was expected that this change would increase the fraction of sequences in the original pool that can bind to the

Zα peptide. In hindsight, the improved protocol for reamplification of the minicircle library would probably have been sufficient for successful *in vitro* selection. A biased library might not be required.

The sequences obtained after five rounds of selection were clearly enriched for those with high propensity to form Z-DNA. The initial rounds of selection had been performed using a large excess of protein over the DNA. By changing the protein/DNA ratio, the stringency of the selection can be controlled. Since during the initial selection cycles the minority of high affinity ligands compete with the far more numerous low affinity ligands, high protein/DNA ratios are useful. By decreasing the protein/DNA ratio only the ligands with high affinity are retained. Therefore, three more selection cycles were done using less protein.

7.3.3. Sequences obtained through minicircle selection

A comparison of the nucleotide composition and the dinucleotide motifs present in the unselected library and the libraries after the 5^{th} and the 8^{th} round of selection is given in Figure 7.1. The selected sequences are depleted in adenosine and thymidine and enriched in cytosine and especially guanosine. The abundance of most dinucleotide motifs is decreased except for CG and GC dinucleotides. Interestingly most other dinucleotides made up of alternating purine and pyrimidine bases are depleted also. The sequences obtained after the 5^{th} and after the 8^{th} round of selection are similar, except that the extent of depletion or enrichment is greater. Figure 7.2 shows graphs in which the frequencies of sequence words of different lengths in the selected sequences have been analyzed. Comparison of the graphs indicates that in sequence words of more than 5 nucleotides length, increasingly no single sequence words are predominant. As in the case of the plasmid DNA selection, it appears that any preferred binding motif has a length of 5 nucleotides or less. Furthermore, there is a clear preference for CG or GC dinucleotides.
The sequence word GGG/CCC is among the most frequent. This is consistent with previous observations that next to CG/GC and GT/CA repeats this sequence requires the least stabilization to adopt the Z-DNA conformation (Ellison et al. 1985). However, in most selected sequences it is found outside the tracts of alternating purines and pyrimidines. The high frequency of its occurrence may reflect the overrepresentation of cytosine and guanosine in the starting library.

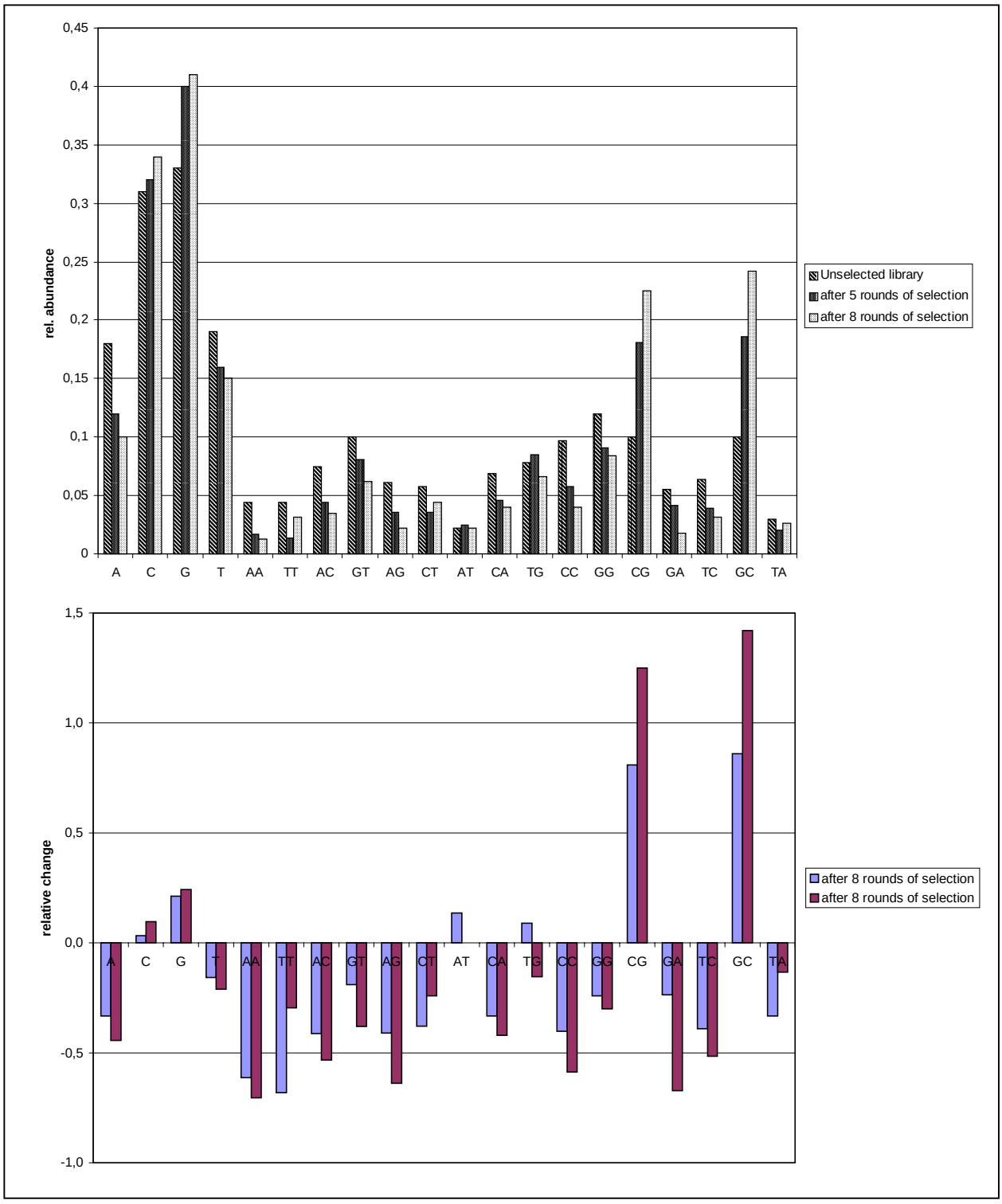
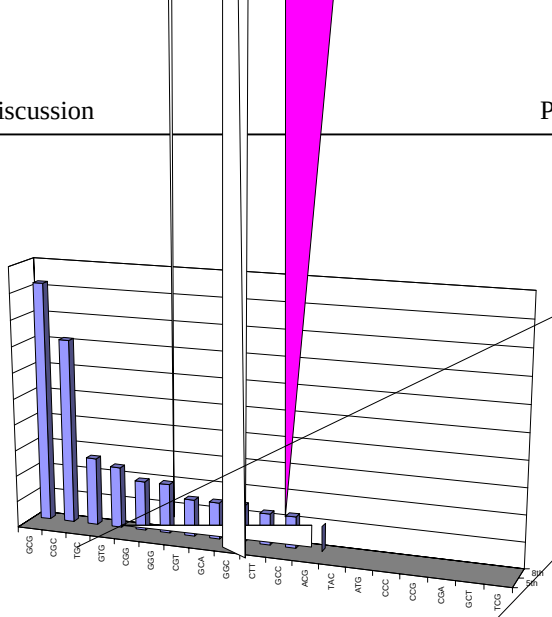


Figure 7.1 Frequencies of mono- and dinucleotides. The abundance of mono- and dinucleotide motifs were determined for the sequences obtained before, and after 5 or 8 rounds of selection. In the upper graph the relative abundance is plotted for each motif. In the lower graph the relative change in their abundances after 5 of 8 rounds of selection is plotted relative to their abundance before selection.



Wordlength : 3

Wordlength : 4

Wordlength : 5

Wordlength : 6

Wordlength : 7

Wordlength 8

Figure 7.2 Relative frequencies of sequence words in the selected minicircles. The sequences obtained before and after 5 or 8 rounds of selection were analyzed for the sequence words contained in them. The frequencies of the sequence words were calculated. The frequencies of the twenty most abundant sequence words obtained after the 8th round of selection are plotted for different sequence words lengths. The frequencies of the same sequence words in the sequences obtained after 5 rounds of selection are included.

The Z-scores of the selected sequences are shown next to the selected sequences in Tables 6.8 and 6.9 . Compared with the Z-scores of the sequences of the unselected library the selected sequences are clearly enriched for those sequences with a high Z-score. Figure 7.3 shows the distribution of the Z-scores of the selected sequences compared to those expected for random sequences (For the source of this graph see Appendix). Besides a cluster of Z-scores around a value of 6 there is a second cluster of Z-scores around a value of 15. The second cluster of higher Z-scores reflects the

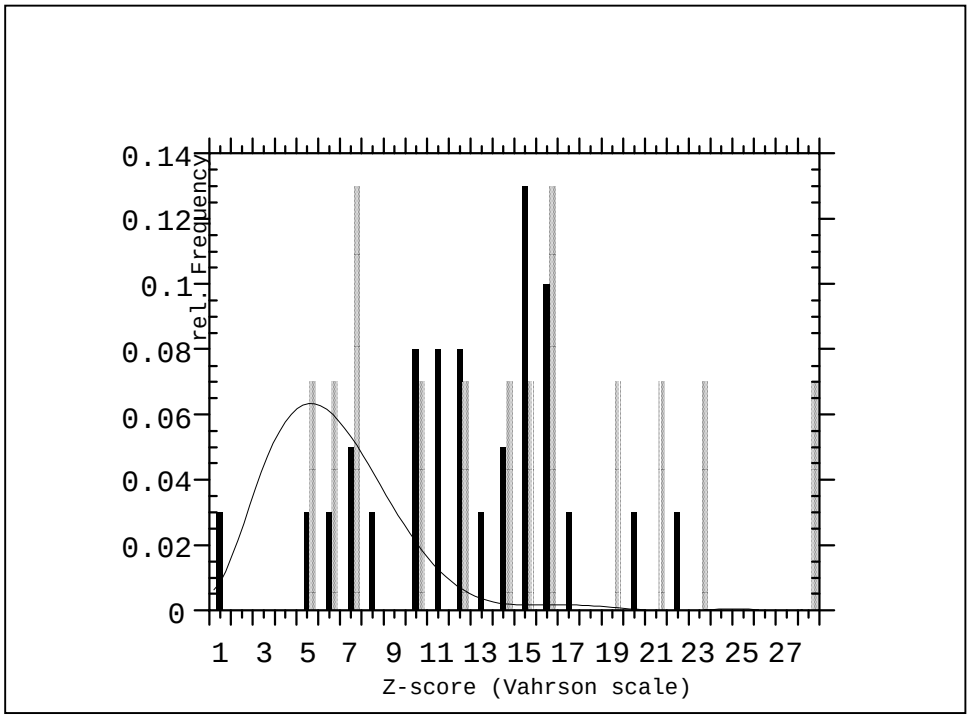


Figure 7.3 The frequencies of Z-scores of the sequences obtained after five rounds of selection are shown in black and those of the sequences obtained after eight rounds of selection are shown in gray. A curve indicating the Z-scores of random sequences has been included for comparison.

selection of sequences with high propensity to form Z-DNA by the Zα peptide. The cluster at lower Z-scores could be interpreted as representing the Z-scores of random sequences. However, often a particular sequence contains two blocks of nucleotides, each with a low Z-score. The Z-scores of the selected sequences are comparable to those of sequences shown to adopt the Z-DNA conformation *in vivo*. The motifs Z1, Z2 and Z3 present in the promoter of the human c-myc gene have Z-scores of 17, 20 / 18 and 30 respectively (Vahrson 1992). However, the motif present in the human corticotropin releasing hormone gene has a Z-score of 124 (Wölfl et al. 1996). A high Z-score does not necessarily mean that this sequence will adopt the Z-DNA conformation *in vivo*. In a study of the human ß-globin gene cluster, the presence of the Z-DNA conformation in 17 sequences with Z-scores between 13 and 38 was analyzed (Müller et al. 1996). Surveying several cell lines no binding of a Z-DNA specific antibody was detected for 10 of these sites. For 5 sequences binding could only be detected in some cell lines and only one sequence appeared to adopt the Z-DNA conformation in all cell lines used. The most likely explanation for these results are different local levels of negative supercoiling.

Closer inspection of the sequences containing several blocks of alternating purine and pyrimidine sequences reveals another important aspect. Depending on the insert the two blocks can be in or out of phase with regard to their sequences of alternating purines and pyrimidines. As shown in Figure 7.4, in the majority of cases the second block of purines

and pyrimidines is in phase with the first block. Furthermore, in those sequences that do not maintain phasing, the blocks of alternation are spaced further apart. This would indicate that interruptions with one or two nucleotides out of phase still allow for the formation of a continuous block of Z-DNA, whereas longer interruptions are inhibitory and

	RY**R**Y**R**Y**R**Y**R**Y**R**Y**R**	
Phasing retained :		# of nucleotides out of phase
G51-6	**G**T**G**C**A**T**A**C**G**C**GAG**C**G**T**G**	1
G51-7	**G**C**G**C**G**C**G**C**GA**CC**G**T**G**C**A**	2
G51-11	**G**C**G**C**G**CCC**G**T**G**T**G**T**G**C**A**	1
G51-27	**G**C**G**C**GGGGG**C**A**C**G**T**G**	3
G51-54	**G**C**A**T**G**CT**GG**C**A**C**G**C**G**C**G**C**G**	2
G51-57	**A**T**G**C**GGG**T**GAG**C**A**C**G**C**G**	2
G51-58	**G**C**G**C**G**C**G**C**GA**TC**G**T**G**T**G**	2
G51-74	**G**C**G**T**GG**C**AG**C**G**C**G**C**G**	2
G81-9	**G**C**G**C**G**CCC**G**C**G**CTT**A**C**G**	1/1
G81-10	**G**C**A**C**G**C**G**C**G**CCT**G**C**G**C**A**	1
G81-22	**G**C**G**C**G**C**G**C**GA**TC**G**T**G**T**G**	2
G81-23	**G**T**G**C**GAG**T**G**T**G**C**G**T**G**C**G**	1
Phasing not retained :		
G51-19	**A**C**A**C**G**TCCC**AG**T**GG**C**G**C**A**	7
G51-60	**G**C**G**T**GG**T**G**C**GGG**C**G**T**G**	2
G81-12	C**G**C**GGGG**T**A**C**A**	2

Figure 7.4 Selected sequences with interruptions in the alternation of purine and pyrimidine bases. Purine bases are shown in bold.

two independent regions of Z-DNA are formed. No particular interrupting sequence seems to be more acceptable than others. In a recent study based on the X-ray crystallographic studies the same result was obtained, the effect of neighboring bases on out of phase base pairs being negligible (Eichman et al. 1999). It is of interest to compare the sequences obtained by in vitro selection with a study, where the Z-DNA binding antibody Z22 was used to isolate genomic sequences from the halophilic bacterium *Halobacterium halobium* (Kim et al 1996). Of the seven sequences published, four sequences contained an interruption in the alternation of purine and pyrimidine bases. Of these, two consisted of the insertion of two nucleotides and the other two had one nucleotide out of phase. Again, the phase of alternation was maintained. Similar results were obtained in the experiments presented here.

7.3.4. Analysis of the selected sequences by bandshift assays and chemical modification

In order to investigate whether the selected sequence can indeed adopt the Z-DNA conformation, two sequences were chosen and analyzed further. First, synthetic oligonucleotides in the form of hairpin oligonucleotides containing the randomized region were used in bandshift assays. Both sequences with alternating purine/pyrimidine bases were bound by GST-Zα peptide. Hence, binding is due to the selected randomized region and not to other features of the minicircle DNA. A third selected sequence had a low Z-DNA score and was chosen as a control. This sequence did not show binding. It appears that the *in vitro* selection process did not yield a population of molecules totally made up of good binders. Some non-binding molecules were retained, probably due to unspecific binding. This is a well known problem of *in vitro* selection experiments.

While the bandshift assays showed the binding activity to be due to the selected sequences and not other aspects of the minicircles, they did not tell much about the conformation of the DNA. Therefore treatment of DNA with diethylpyrocarbonate (DEPC) was used to demonstrate that the selected sequences can indeed adopt the Z-DNA conformation. The rational of the experiment is the ability of certain reactive chemicals to chemically modify DNA if present in the Z-DNA conformation. DEPC reacts with the N7 positions of adenine and guanine. In B-DNA the N7 position is relatively protected from attack. It is much more accessible in Z-DNA. Modification of the DNA by DEPC can therefore be seen as a good indication for the presence of Z-DNA. In order to generate enough material for the analysis, the minicircles were inserted into a general cloning plasmid. Plasmid DNA was then isolated from bacteria, which had the effect of generating negatively supercoiled DNA which in turn stabilized any potential Z-DNA conformation in the plasmid DNA. Again, the same sequences were chosen that had been used for the bandshift assays. Both the sequences containing long stretches of alternating purine and pyrimidine nucleotides were modified upon treatment with DEPC. The extent of the modification was clearly dependent on the incubation time. The control sequence was not modified. These results are a good indication that the selected sequences can indeed adopt the Z-DNA conformation.

Attempts were made to show that the Zα peptide had a higher binding affinity to the selected pool of sequences than to the original one. While PCR reactions done after each selection cycle gave a good indication that this was the case, filter binding assays proved inconclusive (Data not shown). The observed binding constants for both the original library

and the selected pool of sequences were similar. Possibly the reason for this result can be traced to the fact that the majority of DNA molecules present in the minicircle preparation did not represent closed circular DNA. Therefore linear and nicked DNA molecules competed with the minicircles, which may have been sufficient for the binding curve to have been dominated by unspecific binding.

7.4. Is the binding of the Zα peptide to DNA conformation- or sequence dependent ?

The experiments described in this work were prompted by the observation that the Zα peptide binds to DNA oligomers with an oligo (CG) sequence stabilized in the Z-DNA conformation by chemical modification. I therefore set out to obtain more evidence that binding of Zα peptide to DNA is indeed dependent on the Z-DNA conformation. Using circular dichroism spectroscopy it was shown that binding of the Zα peptide to DNA converts the circular dichroism spectrum of the DNA from one typical for the B-DNA conformation into one typical for the Z-DNA conformation. Likewise bandshift assays indicated that those sequences were bound best, which had the highest propensity to adopt the Z-DNA conformation. Therefore it can be concluded that Zα peptide does indeed bind to DNA in a conformation specific way. The second question then became, whether in addition to the dependence on the Z-DNA conformation, any sequence requirement determines the strength of binding. Using bandshift assays it was possible to show that the Zα peptide can not only bind to stretches of alternating CG nucleotides, but that sequences of alternating GT dinucleotides can be bound too, if the sequence is stabilized in the Z-DNA conformation by negative supercoiling. This indicatioed that in principle many, if not all, sequences can bind to the Zα peptide as long as they are present in the Z-DNA conformation.

However this left open the question whether any particular sequence exists which is bound better than others. A bandshift assay in which a $(CG)_4$ sequence was mutated indicated that changes to the sequence resulting in less than five alternating cytosines and guanosines, lead to diminished binding. This, however, could be due to either a true sequence dependence, or to the fact that the changes to the sequence changed their

dinucleotides might be more likely to adopt the Z-DNA conformation than those separated by other dinucleotides. Formation of Z-DNA is a highly cooperative process (Peck and Wang 1983) . The energy required to form the first stretch of Z-DNA is high since it requires the formation of B-Z junctions. Once a stretch of Z-DNA exists, the energy required for its propagation is considerably lower. Formation of the Z-DNA conformation most likely starts at sequences with particular high Z-DNA forming propensity, a condition fulfilled by consecutive CG or GC dinucleotides.

It was therefore decided to use the methodology of *in vitro* selection to investigate the effect of different sequences on the binding of the Zα peptide further. Using either a plasmid library or the minicircle library both times sequences were selected that have a higher propensity to adopt the Z-DNA conformation than random sequences would have. Analysis of the selected sequences showed an accumulation of those rich in runs of CG or GC dinucleotides. The binding of Zα peptide to DNA does not appear to show any sequence specificity other than to those sequences favoring the Z-DNA conformation. Any sequence specificity if it exists is directed to sequence motifs like CGCGC that in itself are among the most likely to adopt the Z-DNA conformation. The current methodology does not allow to distinguish between the two effects. While the data presented in this work cannot exclude a limited degree of sequence specificity, it would have to be small. Any biological effect based on it seems highly unlikely. The conclusion that the binding of the Zα peptide to Z-DNA is dominated by the propensity of a sequence to adopt the Z-DNA conformation is further strengthened by the recently determined x-ray structure of a the Zα peptide bound to DNA. Here, the Zα peptide interacts with the sugar-phosphate backbone of the DNA molecule. The only contact of Za with a base is a van der Waals contact with C8 of guanosine. This contact is not sequence specific since by modeling other bases can be substituted at this position without sterical hindrance (T. Schwartz; personal communication).