6. Biochemical Characterization of the Z-DNA Binding Domain of ADAR1

At the beginning of this thesis project, a domain comprising residues 121-197 of hADAR1 (Herbert *et al.*, 1997), functional in binding Z-DNA, was known. Despite its functionality, the peptide exhibited poor chromatographic behavior and was very unstable upon treatment with small amounts of proteases (data not shown). This indicated that the peptide did not fold into a structurally stable domain. Inhomogeneity in a protein is one of the major reasons for unsuccessful crystallization experiments (McPherson, 1998). Therefore the attempt was made to define a structurally stable, homogeneous domain.

6.1. Cloning of Expression Constructs

As outlined in section 4.1, various expression constructs were cloned. Za131 comprises the functional Zα domain (residues 121-197) plus 25 additional N-terminal and 29 additional C-terminal residues. This construct was used for limited proteolytic experiments to determine a structurally stable Zα core domain. Zab (residues 133-368) spans both homologues motifs Zα and Zβ, including the intervening repeated 49 aa linker module. ZabΔl is identical to Zab, except that it lacks one of the two repeated linker modules (Figure 6.3.5).

6.2. Overproduction of Expression Constructs

His_6-tag fusion proteins were overproduced in *E. coli* and purified under non-denaturing conditions. A fast and efficient purification protocol involving two chromatographic purification steps and removal of the His_6-tag was developed. With this method, described in section 4.17, protein of high purity was obtained within 2 days. The typical yield of pure protein at the end of the purification procedure was 6 mg, starting from 12 mg of crude protein obtained from 1 l bacterial culture.

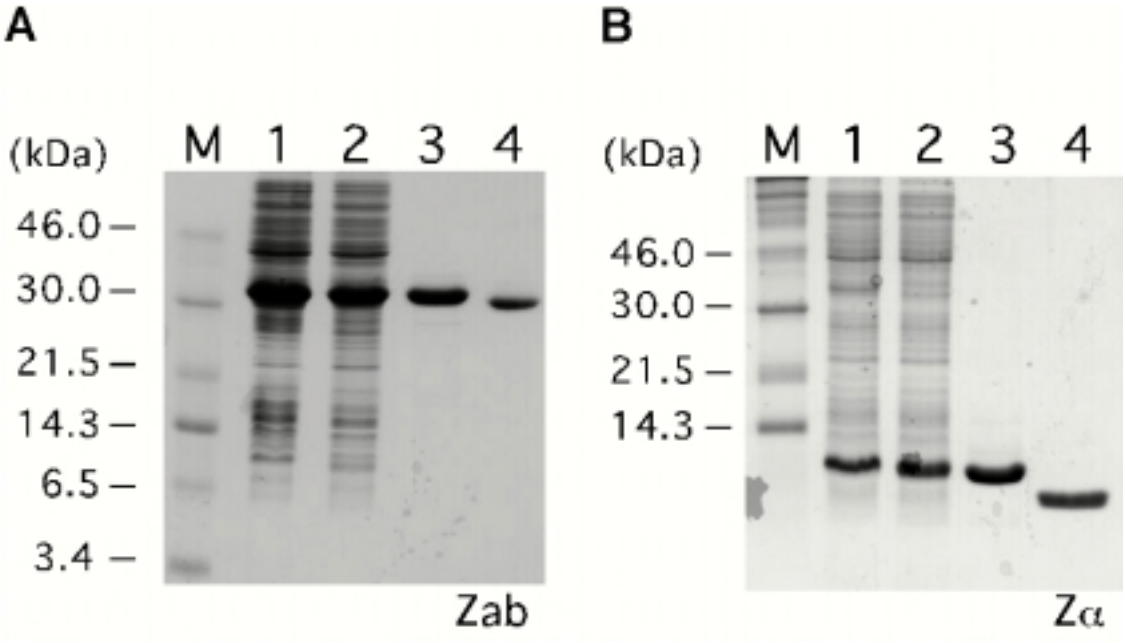

Figure 6.2.1 Protein purification.

Denaturing SDS-PAGE analysis of protein purification steps is shown for Zab (**A**) and Zα (**B**). Lane M, molecular weight markers; lane 1, crude cell extract; lane 2, soluble fraction; lane 3, after Ni-NTA affinity chromatography; lane 4, after removal of the His_6-tag with thrombin and cation exchange chromatography.

6.3. Determination of the Structural Organization of the Bipartite Z-DNA Binding Domain Zab via Controlled Proteolysis

6.3.1. Defining the Boundaries of the Minimal Z-DNA Binding Domain, Zα, of human ADAR1

Limited proteolysis was used to define a structured core containing Zα. The digestion of Za131, comprising residues Gly 96 to Ser 226 with 4 different proteases, endoproteinase Glu-C (V8), chymotrypsin, thermolysin and trypsin, was analyzed. Each of these enzymes has different sequence specificity; therefore, using them in concert results in complementary information. The use of this combination of proteases results in an even distribution of potential cleavage sites throughout the studied protein, with gaps no longer than 6 residues between adjacent sites (Fig. 6.3.1.1).

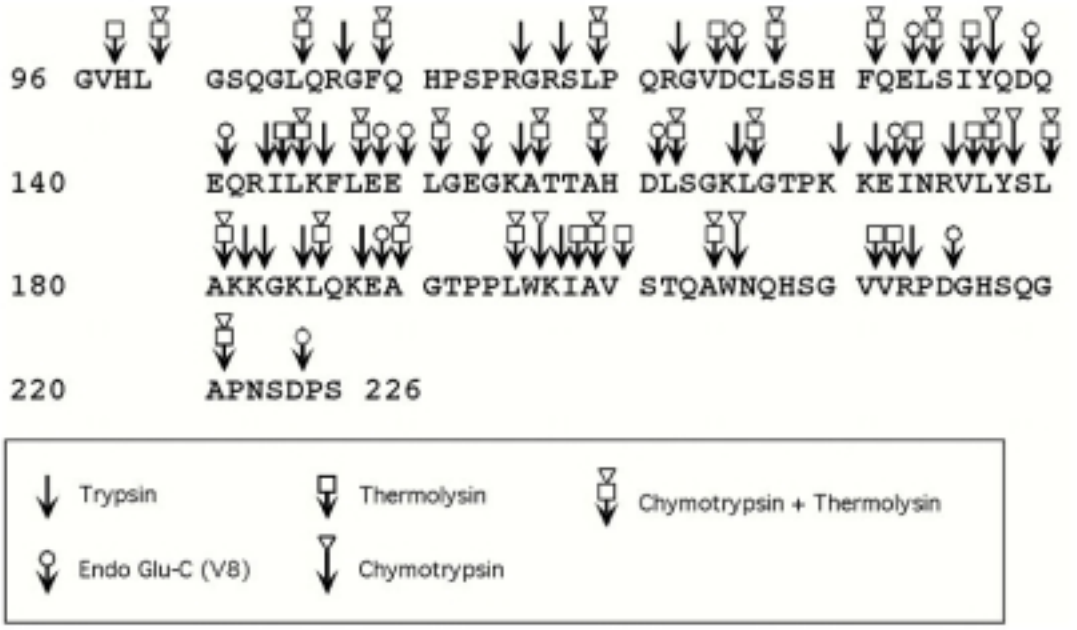

Figure 6.3.1.1 Potential cleavage sites in the Za131 fragment of ADAR1.
Za131 (residues 96-226) was used to define the folded core of Zα, the minimal Z-DNA binding domain in ADAR1. Proteases for the partial digestions were selected in a way to ensure an even distribution of potential cleavage sites throughout the entire sequence. The potential cleavage sites within Za131 of the proteases used are indicated by arrows.

A time course of cleavage with endoproteinase Glu-C (V8) is shown in Fig. 6.3.1.2. An 11 kD band appears rapidly, and increases in intensity over the observed time; the full-length protein band gradually disappears over the same period. The intensity of the 11 kD band is comparable to that of the full-length band, indicating a stoichiometric conversion to a stable product. The cleavage site was mapped to a preferential V8 site, C-terminal to Asp^{132}, using N-terminal sequencing. Similar results were obtained using trypsin and chymotrypsin to cleave this protein (data not shown).

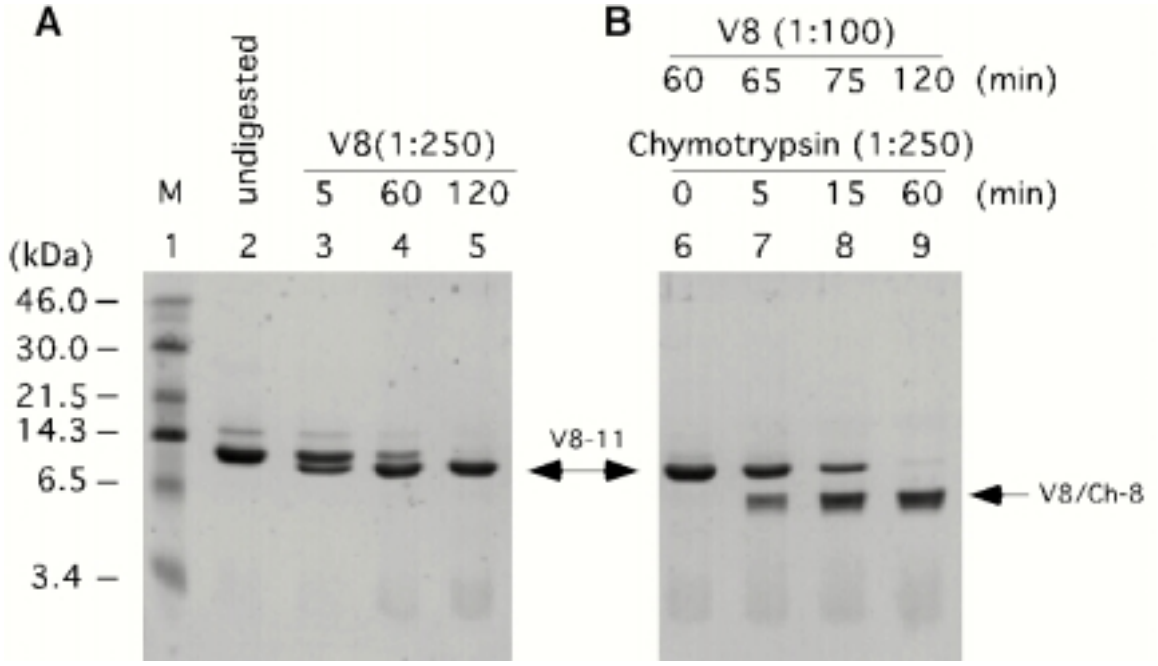

Figure 6.3.1.2 Limited proteolysis reveals a stable *Zα* core domain.
(A) Za131, comprising residues Gly[96] to Ser[226] of hADAR1, was digested with V8 at a protease-to-protein mass ratio of 1:250, at room temperature. Reactions were stopped by heat denaturation after the indicated incubation time. Samples were resolved by SDS-PAGE (18%) and visualized by staining with Coomassie Brilliant Blue. The V8-11 fragment results from a single cleavage and comprises residues 133 to 226. (B**)** The same construct was digested consecutively with two site-specific proteases. After preincubation with V8 (1:100 protease-to-protein) for 1 h, chymotrypsin (Ch, 1:250 protease-to-protein) was added and samples analyzed after the indicated reaction times. V8/Ch-8 is a protease insensitive core fragment containing Zα, which spans residues 133 to 204 of hADAR1.

In order to ensure that a minimum domain had been identified, the protein was cleaved sequentially with two different proteases. Fig. 6.3.1.2, panel B, shows the digestion with V8 followed by chymotrypsin. Before addition of the second protease, only the 11 kD band is detectable. Chymotrypsin further truncates the fragment, producing the stable product, V8/Ch-8. The N- and C-termini of these fragments were identified unambiguously using MALDI-TOF mass spectrometry. The V8-11 fragment was shown to contain residues 133 to 226. Chymotrypsin cuts after Trp[204]; V8/Ch-8 consists of residues 133 to 204. A similar digestion, carried out with V8 and thermolysin, produces a stable product extending from aa 133 to 209 (data not shown). Other combinations of enzymes produced consistent results in all cases; the conclusion is that there is a core domain containing *Zα.* Trp[204] is a potential target for cleavage by both

chymotrypsin and thermolysin. Chymotrypsin cuts well, but thermolysin cuts only marginally at Trp^{204}. Therefore, the core domain was defined as comprising Leu^{133} to Gly^{209}. This core was in no case significantly degraded, while the regions on either end were rapidly degraded to pieces too small to detect.

These results were used to design a stable construct, Zα77, comprising residues Leu^{133} to Gly^{209}. This protein was purified from *E. coli* undegraded under non-denaturing conditions. Zα77 showed superior chromatographic behavior over previous *Zα* constructs, purifying from a Mono S cation exchange column homogeneously, as a sharp peak; this indicated structural uniformity. Samples yield a single band when analyzed on native PAGE (data not shown). When challenged with exogenous proteases, this proteolytically defined Zα domain showed striking stability.

6.3.2. Interaction Between the Two Motifs, Zα and Zβ, to Form a Single Structural Entity

Both *Zα* and *Zβ* are present in every species in which the sequence of ADAR1 is known. The motifs are separated by one or two copies of a module, weakly conserved in sequence, but consistently lacking positively charged residues, and 43-49 aa in length. Twelve residues from this module are an essential part of the stable Zα core domain. It seemed possible that *Zα, Zβ* and the linker module(s) together form a single structural and functional unit. To investigate this possibility, the structural organization of a peptide spanning both DNA-binding motifs was examined. This peptide, termed Zab, and comprising residues Leu^{133} (the previously defined N-terminus of Zα) to Asn^{368} (C-terminal to *Zβ*, from hADAR1), was soluble when overproduced in *E. coli*, and full-length protein could be obtained with high yield. These results indicated proper folding with no significant instability. Improper folding often leads to the formation of inclusion bodies inside the overproducing bacterial cell (Wilkinson & Harrison, 1991). Highly flexible proteins are frequently degraded if expressed in a foreign host (Makrides, 1996).

Fig. 6.3.2.1 shows the results of the digestion of Zab with 4 different proteases.

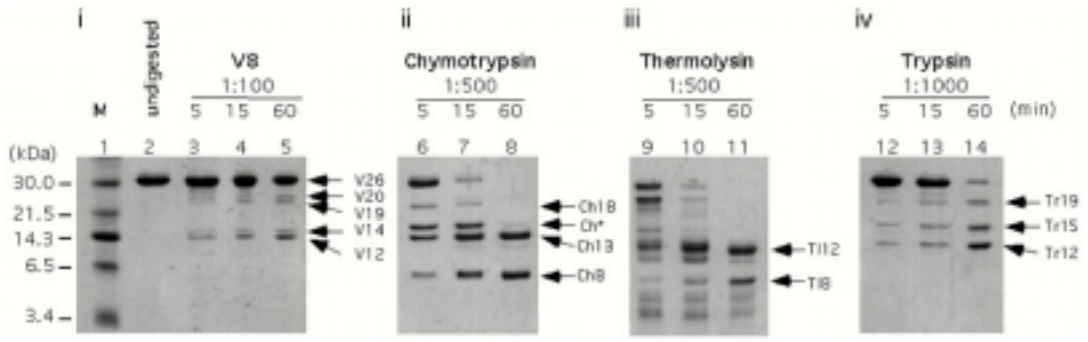

Figure 6.3.2.1 Digestion of Zab with different proteases.
Zab (residues 133-368 from hADAR1) was incubated with V8 (i), chymotrypsin (ii), thermolysin (iii) and trypsin (iv), at the indicated protease-to-protein mass ratios for the indicated times. Fragments were resolved by SDS-PAGE (18%) and visualized by staining with Coomassie Brilliant Blue. Arrows indicate stable fragments, which were further analyzed by mass spectrometry. The results of this analysis are shown in Table 6.3.2.2.

Each enzyme cleaves in a characteristic pattern, and produces a small number of very stable bands. Time points were selected to allow the identification of all stable products, using mass spectrometry and additionally N-terminal sequencing when necessary; minor products were identified wherever possible. In each case well-resolved spectra were recorded. Table 6.3.2.2 lists the peptides produced by each enzyme, as determined from their molecular weights. For chymotrypsin, trypsin and V8 the assignments are unambiguous and in good agreement with SDS-PAGE analysis. Minor exceptions are fragments Tr 8 and Ch 5, which were detected only by mass spectrometry, as discussed below. In the case of thermolysin, it was not possible to unambiguously assign the multiple transitory fragments; however, the major fragments seen after 60 min of digestion can be identified.

Protease	Proteolytic fragment	Measured mass	Sequence assignment	Calculated mass
		Da		Da
Endo Glu-C (V8)				
	V26	25,650	$L^{133} \ldots E^{361}$	25,666
	V20	19,756	$L^{133} \ldots L^{307}$	19,716
	V19	18,841	$L^{133} \ldots K^{300}$	18,850
	V14	13,536	$D^{240} \ldots E^{361}$	13,539
	V12	12,173	$L^{133} \ldots E^{239}$	12,126
Chymotrypsin				
	Ch18	18,082	$N^{205} \ldots N^{368}$	18,119
	Ch13	12,918	$N^{254} \ldots N^{368}$	12,907
	Ch8	7,865	$I^{135} \ldots W^{204}$	7,851
	Ch5	5,189	$N^{205} \ldots W^{253}$	5,230
Thermolysin				
	Tl12	12,239	$R^{261} \ldots M^{363}$	12,186
	Tl8	8,495	$S^{134} \ldots G^{209}$	8,460
Trypsin				
	Tr19	19,103	$L^{133} \ldots K^{302}$	19,102
	Tr15	15,309	$N^{233} \ldots N^{368}$	15,310
	Tr12	11,600	$L^{133} \ldots R^{232}$	11,587
	Tr8	7,559	$N^{233} \ldots K^{302}$	7,539

Table 6.3.2.2 Mass spectrometric analysis of Zab fragments.

A schematic diagram of the major transitory products and the stable proteolytic fragments is shown in Figure 6.3.2.3.

V8 cleaves Zab rapidly at a single site, Glu^{361} at the extreme C-terminus (Fig. 6.3.2.1, i). The resulting peptide is very stable to further proteolysis, in spite of an abundance of potential cleavage sites, including Asp 132, which is exquisitely sensitive in the shorter construct used to define Zα, as described above. After long incubation with high amounts of enzyme, additional cleavage occurs at the sites Glu^{239}, Glu^{301} and Leu^{307}. The Glu^{239} site lies within the first 49-aa repeat; remarkably, the equivalent site in the second repeat,

Chymotrypsin cleaves the protein after Trp^{204} and Trp^{253}. These sites are at equivalent positions in the two copies of the tandem repeat. The three generated fragments are stable (Fig. 6.3.2.1, ii). The 5 kD fragment is not visible on SDS-PAGE, although it generates a signal in mass spectrometry of comparable intensity to Ch 12 and Ch 8. Coomassie staining depends largely on positive charges present in the peptide (Tal *et al.*, 1985). The 49 aa repeat contains only one positively charged residue. Therefore, the Ch-5 fragment might be resolved on the gel, however it is not stained. Two other transitory fragments, Ch 18 and Ch∗, were separated in the gel. Ch 18 could be assigned to be the product of a single cutting site. Ch∗ could not be unambiguously determined by means of mass spectrometry.

Thermolysin produces similar stable products (Fig. 6.3.2.1, iii). Again, symmetrical sites in the repeated linker, Gly^{209} and Gly^{258}, are cut, resulting in two stable products on an SDS-PAGE. Because thermolysin has low sequence specificity, many transitory products are seen, especially at early time points. Because of these products, it was not possible to unambiguously identify the Tl 5 fragment from among several candidates seen by mass spectrometry.

Trypsin attacks the protein at two preferred sites: Arg^{232} in the first repeat, and Lys^{302} near the N-terminus of $Z\beta$ (Fig. 6.3.2.1, iv). (The site equivalent to Arg^{232} in the second repeat is Ser^{280}, not a substrate for trypsin.) Two sites near the C-terminus, Lys^{366} and Arg^{367}, result in heterogeneity of the full-length protein and in the Tr 15 fragment. Most of the expected products are stable; however, the C-terminal region peptide, starting at Ile^{303}, is detected neither on SDS-PAGE nor by mass spectrometry. A similar result is seen after extensive V8 digestion — again the C-terminal fragment is not stable. It appears that $Z\beta$, intrinsically more accessible than $Z\alpha$, is stable only in the context of the larger domain.

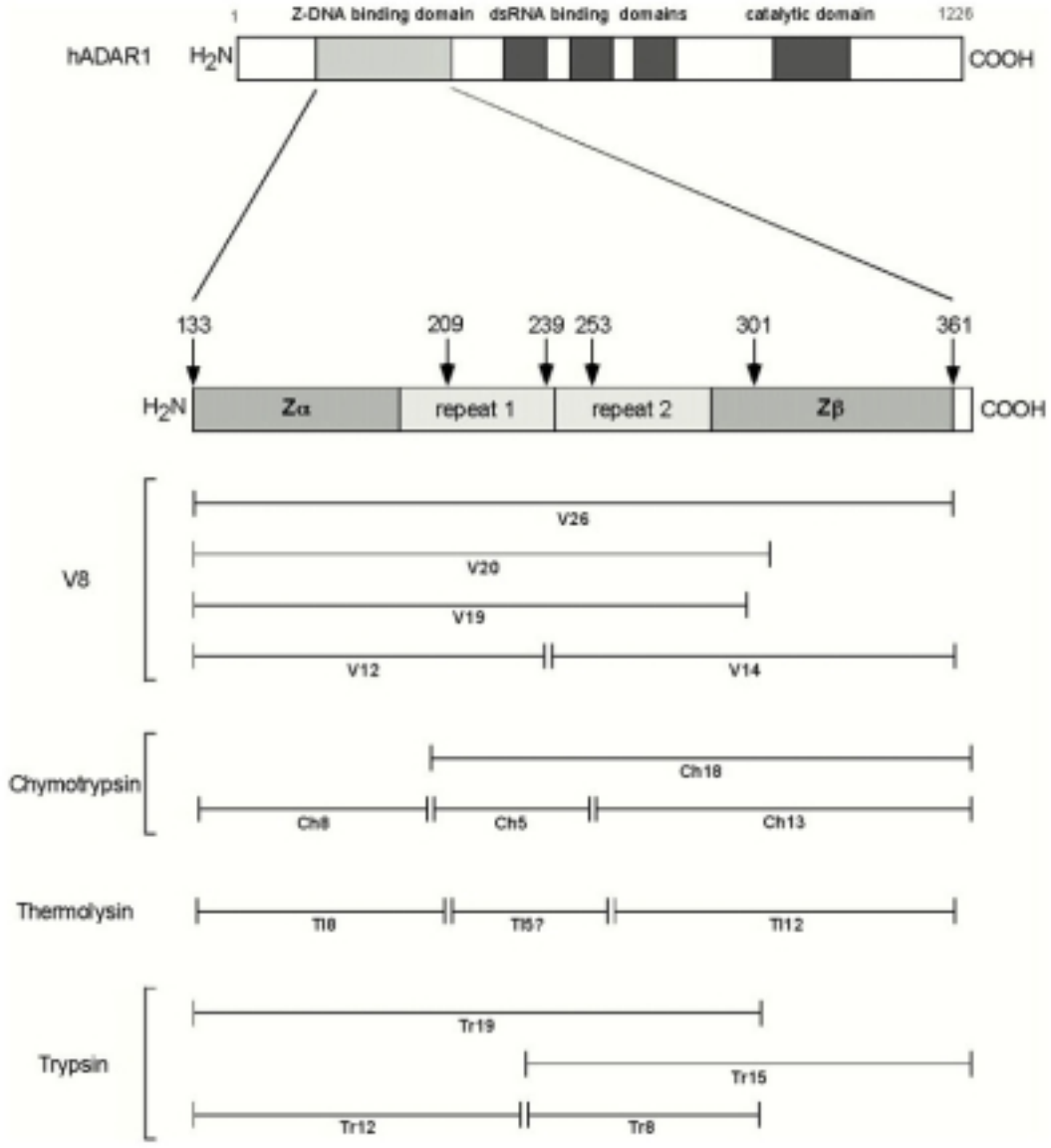

Figure 6.3.2.3 Structure and protease cleavage map of the Z-DNA binding domain of ADAR1.
At the top is a schematic representation of human ADAR1. Below are the stable fragments produced by limited proteolysis. Numbers above ADAR1 are residue positions. The illustrations are proportional.

6.3.3. The Role of the Linker Region for Stabilizing of the Zab Domain

In all the ADAR1 genes sequenced, there is a single copy of a 43-49 aa linker module between $Z\alpha$ and $Z\beta$.In hADAR1 this module is repeated. To determine the effect of this repeat on the structure of Zab, a protein lacking one module was constructed. This protein, ZabΔl, was produced in high yields as a soluble protein in *E. coli*, and could be purified to homogenity. Protease mapping shows results similar to those for Zab (data not shown). Trypsin and V8 cleave at the identical residues. Chymotrypsin and thermolysin have only a single site each. Therefore, the overall structure of the domain is not altered by the presence of the repeated module.

6.3.4. Conclusions from Secondary Structure Examinations of Zab

Circular dichroic measurements in the region between 190 and 250 nm are a useful tool to assess the secondary structure of a protein. This method was used to analyze Zα, Zab, and ZabΔl. The spectra are shown in Figure 6.3.4.1, along with a difference spectrum between Zab and ZabΔl, which reflects the contribution of a single copy of the linker module. Results of the analysis of the curves using the program K2d (Andrade *et al.*, 1993) are shown in Table 6.3.4.2. The $Z\alpha$ and $Z\beta$ motifs contain significant amounts of α-helix and β-sheet structures. In contrast, to a large extent the linker adopts an alternate structure. This is consistent with the secondary structure analysis of the amino-acid sequence with computer programs, such as PHD (Rost & Sander, 1994). Those analyses predict that no significant areas in the linker are structured as α-helices or β-sheets. It must be emphasized that the proteolysis studies clearly indicate that the linker module is structurally well-defined, although in the majority neither α-helical nor β-pleated.

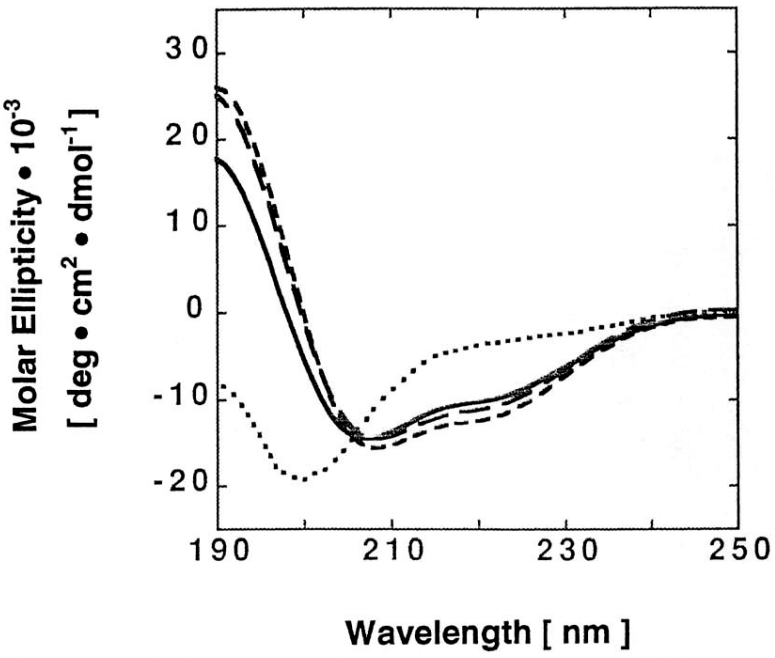

Figure 6.3.4.1 Protein CD spectra.

Spectra are expressed in terms of mean residue ellipticity in units of deg cm^2 $dmol^{-1}$. The curves show the protein spectra of Zab (———), ZabΔl (— — — —) and Zα (- - - - - -). A difference spectrum, Zab minus ZabΔl, is also shown (·····). The corresponding percentages of secondary structure motifs were calculated using the program K2d (Andrade *et al.*, 1993) and are listed in Table 6.3.4.2.

Protein	α-Helix	β-Sheet	Other
	%	%	%
Zα	40	16	44
ZabΔl	37	16	47
Zab	29	16	55
49 aa repeat	8	27	65

Table 6.3.4.2 Secondary structure analysis of recorded CD spectra (Figure 6.3.4.1).

6.3.5. Proteolytic Sensitivity of Zab in Presence and Absence of Ligand

The presence of substrate can affect the protease sensitivity of a protein either directly, by sterical interference of the substrate molecule, or indirectly, by altering the conformation of the protein. In order to test whether this is the case for Zab, protease digestions were carried out in the presence of either B-DNA or Z-DNA. Although there were no dramatic changes in the digestion profiles, B-DNA stabilizes Zab slightly against proteolysis, and Z-DNA has a very marked stabilizing effect.

Digestion experiments with V8 are shown in Fig. 6.3.5.1. Although there is no protection of the C-terminal site (Glu^{361}), the internal cleavage sites are strongly protected in the presence of Z-DNA. Cleavage sites at residues Glu^{301} and Leu^{307} are completely protected, resulting in the absence of the V20 and V19 bands. Cleavage at residue Glu^{239} is reduced, with an approximately 50-fold increase in the stability of the full length Zab protein relative to the absence of DNA. In contrast, B-DNA protects Zab against cleavage only approximately 5-fold, and does not alter the choice of sites.

In additional experiments, chymotrypsin, thermolysin and trypsin all cut at their established sites, but to an approximately 50-fold lower extent in the presence of Z-DNA (data not shown).

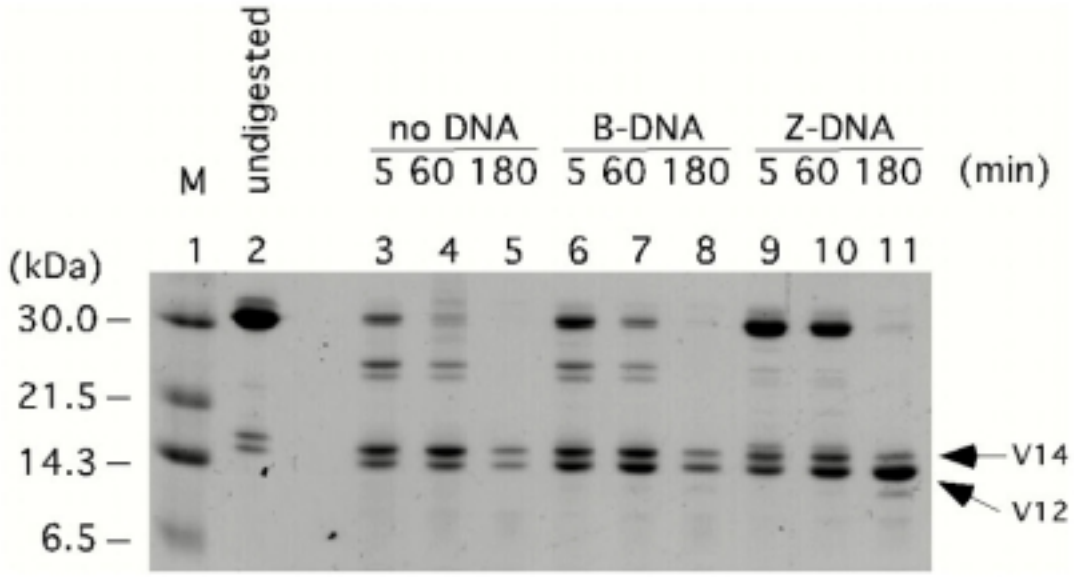

Figure 6.3.5.1 Zab is protected from proteolysis in the presence of Z-DNA.
Zab was digested with V8 without DNA (lanes 3 – 5), in presence of B-DNA (lanes 6 – 8) and in the presence of Z-DNA (poly d($^{5\text{-Me}}$C-G)) (lanes 9 – 11). The protein-to-protease was ratio 1:30.

These results suggest that the entire domain becomes more rigid and less accessible in the presence of ligand. The protection of sites within $Z\beta$ from V8 cleavage may occur because these sites are involved in DNA interaction. On the other hand, conformational changes occurring in the protein as a consequence of binding to DNA could prevent V8 from cutting. It is of note that the nearby trypsin site, Lys^{302}, is protected in the presence of Z-DNA, but not to the same extent (data not shown).

When Zab in the presence and absence of Z-DNA is compared, there is no change in the CD spectra between 190 and 250 nm (data not shown). This indicates that there is no major change in the secondary structure of the protein when substrate is bound.

6.4. Z-DNA Binding Behavior of Zab and Subdomains

The binding of *Zα* to Z-DNA has previously been characterized using electrophoretic mobility shift assays (Herbert *et al.*, 1993, 1995, 1997). This assay was used here to compare the binding of Zα to that of Zab. $d(^{5\text{-Br}}CG)_{20}$ was used as a substrate; this oligonucleotide is stabilized in the Z-form by the presence of bromine in the 5-position of cytosine (Malfoy *et al.*, 1982). Binding was tested in the presence of a 10^4-fold excess of B-DNA. The results are shown in Figure 6.4.1. Four different proteins were compared at four concentrations (500 nM, 100 nM, 20 nM and 4 nM protein). Zab binds well at 500 nM and 100 nM, producing a stable, high-molecular-weight protein-DNA complex (Fig. 6.4.1 a). At lower concentrations, the complex appears to break down during electrophoresis, resulting in a smear. This behavior suggests that the most stable complex is formed when the sites on the probe are saturated. ZabΔl shows a similar behavior, although the stable complex is formed only at the highest concentration (Fig. 6.4.1 d). In contrast, Zα produces two complex bands, which appear smeared at all concentrations (Fig. 6.4.1 b). Compared to Zab, Zα has a slightly higher apparent affinity for the substrate. The smearing is the result of the instability of the complex under electrophoretic conditions and the longer migration path of Zα-DNA-complexes as compared to Zab-DNA-complexes.

These results may indicate that binding of the Zβ moiety of Zab is responsible for the difference in binding behavior between Zα and Zab. To test this hypothesis, Zab was digested with chymotrypsin and then assayed in the bandshift (Fig. 6.4.1 c). Complete digestion, yielding the *Zα* and *Zβ* motifs as separate peptides, was confirmed by SDS-PAGE (Fig. 6.4.1 panel B). This mixture shows a bandshift pattern very similar to that for Zα. No additional bands are observed; therefore *Zβ* alone does not bind to the substrate. Since the molecular weights of the *Zα* and *Zβ* containing fragments differ substantially, it is extremely unlikely that any complex formed by *Zβ* and DNA would comigrate with the observed *Zα*-DNA complexes (Lane *et al.*, 1992). That the isolated *Zβ* is not capable of binding Z-DNA under these conditions is remarkable considering the conservative substitution of functionally important

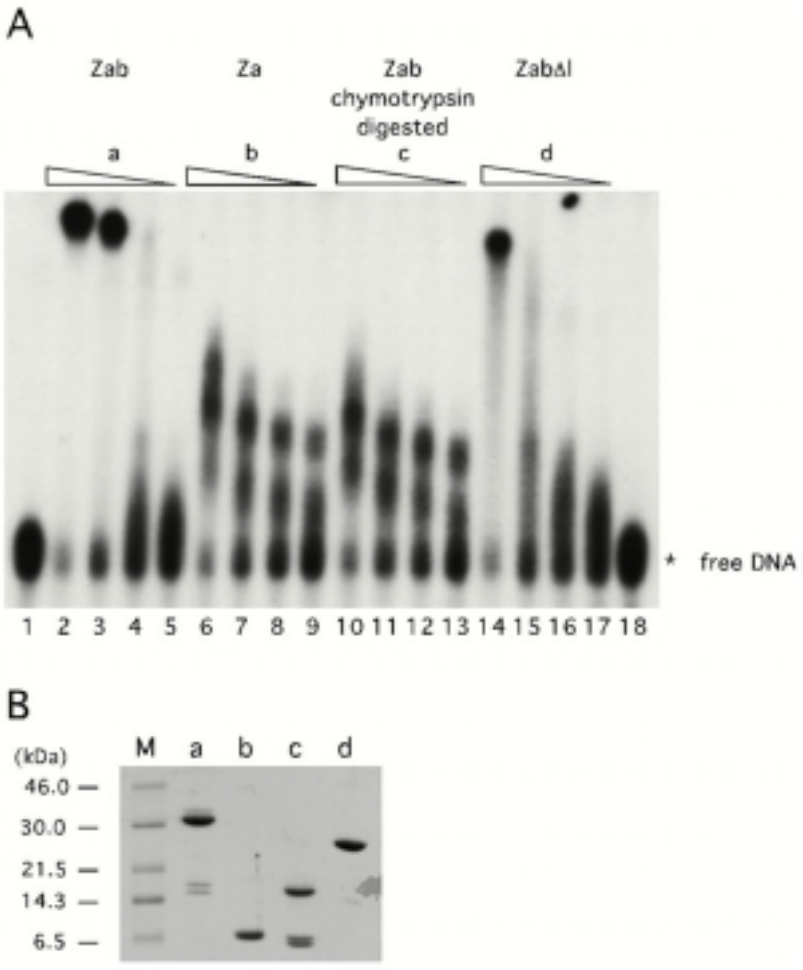

Figure 6.4.1 Binding of Z-DNA by Zab and subdomains.

(**A**) Electrophoretic mobility shift assays were performed with ^{32}P-d($^{5\text{-Br}}$CG)$_{20}$ as the ligand, which is stable in the Z-conformation under the applied conditions (Malfoy *et al.*, 1982). Zab (a, lanes 2-5), Zα (b, 6-9), Zab digested with chymotrypsin to separate the $Z\alpha$ and $Z\beta$ motifs (c, 10-13) and ZabΔl (d, 14-17) were each assayed in a 5-fold dilution series (500 nM, 100 nM, 20 nM, 4 nM). Lanes 1 and 18 show the migration of free ligand. The spot at the top of lane 16 is an artifact, which was not reproducible. (**B**) SDS-PAGE (18%) of 200 nmol of the protein preparation used for the bandshift assay in panel A. The bands were visualized by Coomassie Brilliant Blue staining. Proteins are labeled as in panel A. The digestion of Zab with chymotrypsin (c) is complete, leaving no full-length protein. The difference in the size of Zα and the small digestion fragment in lane c is due to 5 additional C-terminal residues present in the Zα expression construct.

As a second method of studying the binding of Z α to DNA, circular dichroism was used to monitor the transition of the DNA conformation from the B-form

to the Z-form (Herbert *et al.*, 1997, 1998; Berger *et al.*, 1998). The spectrum of Z-DNA is inverted as compared to that of B-DNA in the near UV region between 240 nm and 300 nm (Pohl & Jovin, 1972; Riazance-Lawrence & Johnson, 1992). Figure 6.4.2 shows the spectra of two Z-DNA-forming oligomers of different sequence, $d(CG)_6$ and $d(CA)_7 \bullet d(TG)_7$, in the presence of either Zα or Zab. The DNAs adopt the right-handed B-DNA conformation in the absence of protein. Protein was added in aliquots, resulting in protein-to-base pair molar ratios of 1:6, 1:4, 1:2 and 1:1.5. When Zα is added, the spectra of both $d(CG)_6$ and $d(CA)_7 \bullet d(TG)_7$ become inverted, indicating the shift from the B- to the Z-DNA conformation, with saturation at a 1:2 ratio (Fig. 6.4.2 panel C, D). Further addition of protein does not change the spectrum significantly above 250 nm, where the contribution of the protein to the spectrum is negligible. Below 250 nm, the spectrum is dominated by the contribution of the protein. This result is in agreement with Herbert *et al.* (1998), which showed that a *Zα* motif (aa 121-201) binds without sequence specificity to a variety of Z-DNA forming sequences. Zab converts $d(CG)_6$ to the Z-form with a similar stoichiometry to Zα (Fig. 6.4.2 panel A). In contrast, there is only a limited effect on $d(CA)_7 \bullet d(TG)_7$ with the addition of Zab, even at a ratio of 1:1.5 (Fig. 6.4.2 panel B).

Zα is able to bind to Z-DNA in a sequence-independent manner. However, when *Zα* is in the context of the entire bipartite domain, Zab, a sequence preference for $d(CG)_n$ is observed. Bandshift data suggests that the mode of binding is different between Zα and Zab, probably reflecting a difference in the degree of cooperativity.

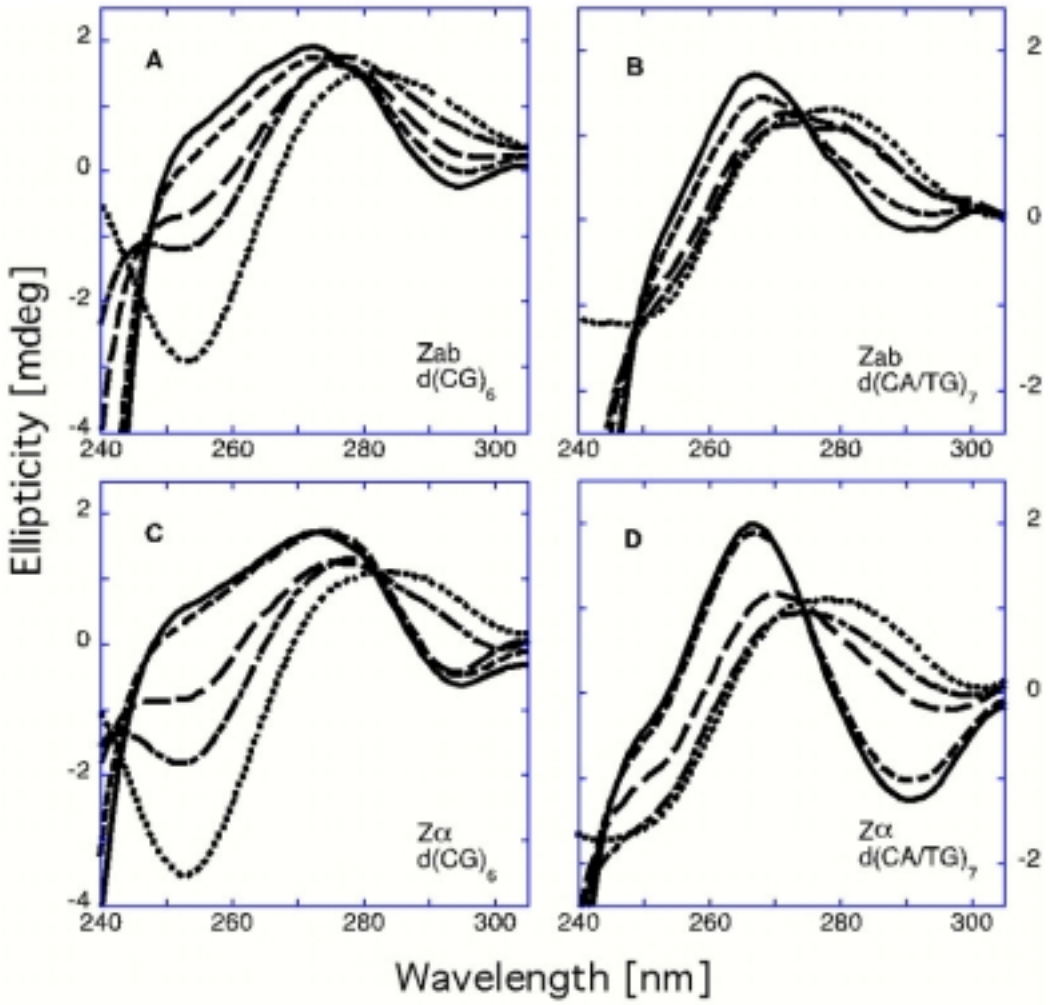

Figure 6.4.2 CD studies of the conformational change of Z-forming DNA sequences in the presence of Zα or Zab.

The spectra show the titration of $d(CG)_6$ (panels A, C) and $d(CA)_7 \bullet d(TG)_7$ (panels B, D) with Zab (A, B) and Zα (C, D), respectively. The curves represent the spectra of the DNA alone (————), and in presence of protein at a protein to base pair molar ratio of 1:6 (– – – – – –), 1:4 (— — — —), 1:2 (— · — · —) and 1:1.5 (·····),respectively. Spectra are expressed in absolute values of ellipticity in millidegrees.