8. Discussion

8.1. The Z-DNA Binding Mode of Zα

This crystal structure reveals that Zα is tailored to recognize Z-DNA in a conformation-specific manner. The protein exhibits a surface complementary in shape and electrostatic nature to Z-DNA. That the interaction with Z-DNA is observed in three different packing environments is a convincing result of this study. It underscores the likelihood that this interaction would also be observed in solution. *In vitro* experiments on Z-DNA binding by Zα are largely in agreement and can be explained with this complex crystal structure. Scanning alanine mutagenesis experiments suggested that the side chains of Lys^{169}, Asn^{174} and Tyr^{177} are involved in crucial DNA contacts. DNA binding was largely diminished by alanine mutations. Pro^{192} and Trp^{195} were found to be important for both structural integrity and DNA binding of Zα (Schade *et al.*, 1999). These findings are in excellent agreement with the crystal structure. Mutation of the other residues involved in DNA contacts, as shown in this study, did not have similarly strong effects. Changing Lys^{170}, Thr^{191} or Pro^{193} to alanine, one at a time, appeared tolerable. An open question is, whether double mutations would have stronger effects on DNA binding.

It is interesting to compare the degree of conservation among Zα sequences from different species (Fig. 8.1.1). The residues involved in DNA contacts are, with one exception, strictly conserved between all species. This means that the conservation of contact residues is even higher than the conservation of residues contributing to the hydrophobic core. Thr^{191} is the only residue that is mutated to serine (bovine) or lysine (mouse, rat). With respect to the results obtained from scanning alanine mutagenesis studies, these alterations at Thr^{191} are expected to be tolerable for effective Z-DNA binding.

In this crystal structure, two Zα monomers bind the 6 bp DNA substrate in a symmetrical fashion. There is no direct or indirect contact between the two domains. It appears that the symmetrical positioning observed is due to the chosen DNA oligomer and favorable crystal packing interactions, and does not necessarily reflect the orientation that might occur in a living cell. Various *in vitro* experiments gave rise to the speculation, whether Zα might bind as a dimer to Z-DNA (Herbert *et al.*, 1998). This crystal structure does not support

this hypothesis. True dimerization, involving a dimer interface between the two monomers, is not seen here and it is therefore very unlikely that it would play a decisive role under different conditions. Nevertheless, there are reasons to speculate that one Zα monomer is not sufficient to form a stable complex with Z-DNA under non-crystalline conditions. The solvent-accessible surface area on Zα is diminished by 430 $Å^2$ upon binding to Z-DNA (calculated with X-PLOR). The loss of solvent-accessible surface area contributes significantly to the free energy of binding (Lilley, 1995). The area lost upon DNA binding in the case of Zα is remarkably low — less than half of the area lost in typical other protein-DNA interfaces (Lilley, 1995). This result might explain the instability of the Zα-DNA complex observed in gel retardation assays (Fig. 6.4.1). Zab, on the other hand, shows a very stable complex formation with Z-DNA. Given the sequence similarity between Zα and Zβ, it is attractive to consider both domains to bind in concert as a single bipartite domain. This would effectively double the loss of solvent-accessible surface area, even under the assumption that the intervening linker region would not contact DNA at all. It must be pointed out that Zα and Zβ, although they likely have a very similar tertiary structure, presumably do not bind DNA in an identical way. Many indications support this hypothesis. The alignment of the Zα and Zβ primary sequences, based on the Zα structure determined in this study, reveals three marked differences between Zα and Zβ (Fig. 8.2.1). First, helix α1 appears to be roughly one turn shorter in Zβ. This should not change the overall fold, since the first turn of α1 in Zα is completely solvent-exposed. It does not contribute to the hydrophobic core, nor does it interact in another significant way with other parts of the domain. Second, Zβ has one residue inserted at the beginning of helix α3. It is likely that this architectural change affects the binding interface, although in which way is difficult to predict. Third, the position of Tyr^{177} in Zα is occupied by an isoleucine or asparagine in Zβ. Therefore, the network of crucial interactions involving Tyr^{177} in Zα must be perturbed in Zβ. The fact that the isolated Zβ domain, in contrast to Zα, has a negligible affinity for Z-DNA *in vitro*, also indicates that Zα and Zβ will have different orientations on the DNA. This is further supported by the result presented in this study that Zab does not exhibit the sequence-non-specific binding behavior observed for the isolated Zα domain. Recently there have been numerous examples, where two HTH proteins are paired or form heterodimers to act in concert (Passner *et*

al., 1999; Piper *et al.*, 1999; Xu *et al.*, 1999; Zheng *et al.*, 1999). It is an evolving idea that nature uses this combinatorial approach efficiently to accomplish different ligand specificities.

8.2. Z-DNA *versus* B-DNA Binding by HTH Proteins

HTH proteins are one of the largest and most intensively studied protein families to date. First, the HTH motif has been characterized in bacterial gene regulatory proteins (reviewed in Harrison & Aggarwal, 1990, Pabo & Sauer 1992). The conventional designation of the HTH (Pabo & Sauer, 1984) is a twenty-residue segment corresponding to two helices and an intervening turn region found in the first three HTH structures, which fold into an elbow shape. The HTH is properly described as a 'motif' because it is a recurring substructure seen in otherwise different, cooperatively folding domains, and because it is not in general a stably folded unit on its own (Harrison & Aggarwal, 1990). The structures of a large number of HTH proteins are now available. Many of those proteins, from prokaryotic and eukaryotic origin, deviate to various extents from the original fold. Attempts to classify those deviations have been made on functional and structural grounds. Structurally-based classifications emphasize common features of the domain fold in which the HTH motif occurs (Wintjens & Rooman, 1996). Based on this approach, Zα belongs to the family of HTH proteins with α/β topology (Fig. 7.10.2). HTH proteins are the best-studied class of DNA-binding proteins, and the canonical interaction with right-handed B-DNA is well characterized. Typically, the second helix of the HTH motif, called the recognition helix, is cradled in the major groove. There, the base edges expose a characteristic pattern of potential hydrogen bond donors and acceptors, which allow for base-specific recognition (Seeman *et al.*, 1976). In the family of HTH proteins with α/β topology, DNA contacts are also often made between residues in the N-terminal region of the first helix of the HTH motif and the phosphate backbone. In variants of the family, additional contacts extend the interaction surface (Fig. 8.2.1).

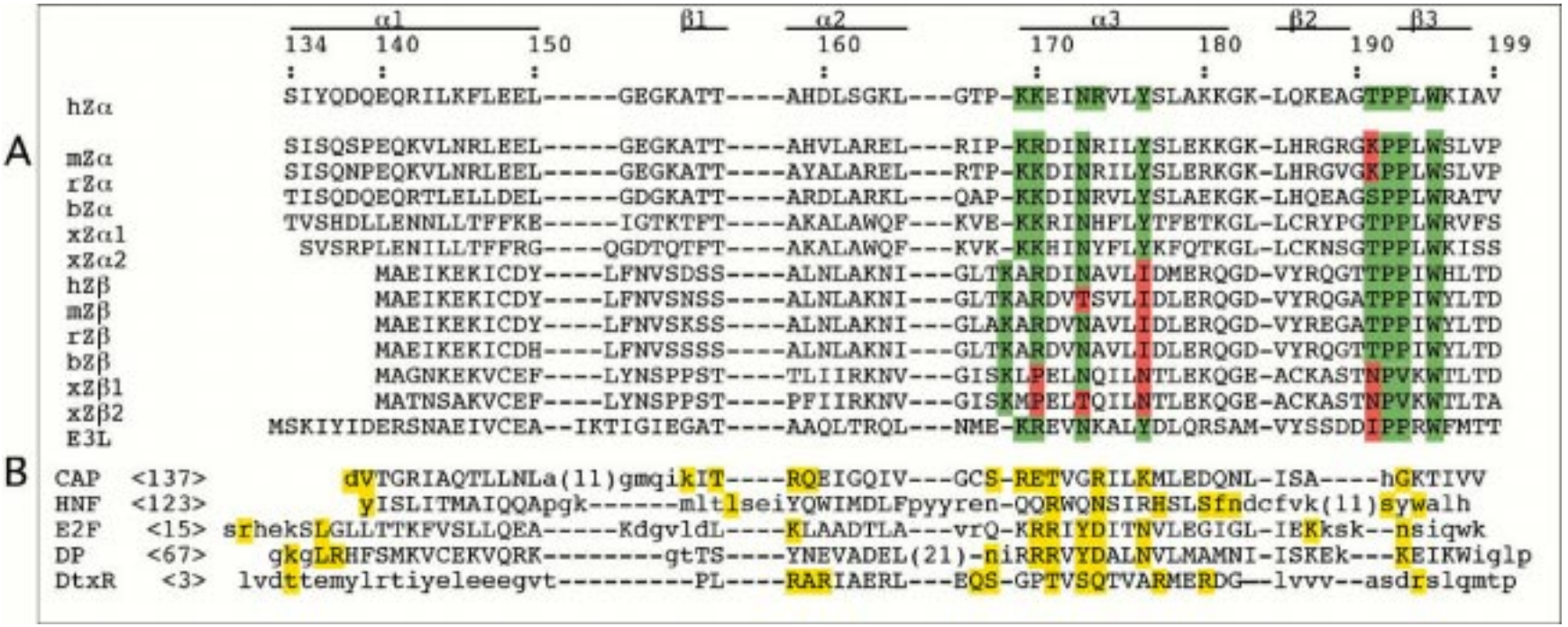

Figure 8.2.1 Sequence alignment.

Numbering and secondary structural elements for Zα from human ADAR1 (hZα) are indicated above the sequences. Residues that contact Z-DNA in hZα are green. (**A**) Sequence alignment of Zα homologs. Residues are green if they are identical or chemically similar to DNA contacts observed in hZα, otherwise they are red. The sequences are: hZα, hZβ, ADAR1 *Homo sapiens*, accession number (acc) GenBank U18121; mZα, mZβ, ADAR1 *Mus musculus*, acc GenBank AF052506; rZα, rZβ, ADAR1 *Rattus norvegicus*, acc GenBank U18942; bZα, bZβ, ADAR1 *Bos taurus*, (5); xZα1, xZβ1, dsRAD-1 *Xenopus laevis*, acc GenBank U88065; xZα2, xZβ2, dsRAD-2 *X. laevis*, GenBank U88066; E3L *Vaccinia virus* Western Reserve strain, acc GenBank S64006.
(**B**) Structurally related B-DNA binding HTH proteins.are shown. They were aligned with the program Superimpose (M. A. Rould, personal communication). Residues in structurally equivalent positions to Zα are in uppercase letters. DNA contacts with B-DNA (taken from published complex crystal structures) are in yellow. The sequences are: Catabolite gene activator protein CAP, acc Protein Data Bank (PDB) 1CGP; H5; hepatocyte nuclear factor HNF-3γ, acc Nucleic Acid Database PDT013; transcription factors E2F and DP, acc PDB 1CF7; diphteria toxin repressor DtxR, acc PDB 2DTX.

One example is illustrated by the structure of HNF-3γ bound to B-DNA (Fig. 8.2.2, panel B), where contacts include residues in two extended loops, whose importance and spatial arrangement gave rise to the name 'winged-helix' motif. The mode of interaction of Zα with DNA (Fig. 8.2.2, panel A) is markedly different from that of B-DNA-binding HTH proteins, reflecting the fact that Z-DNA, unlike B-DNA, has no major groove and a very deep, nearly inaccessible minor groove.

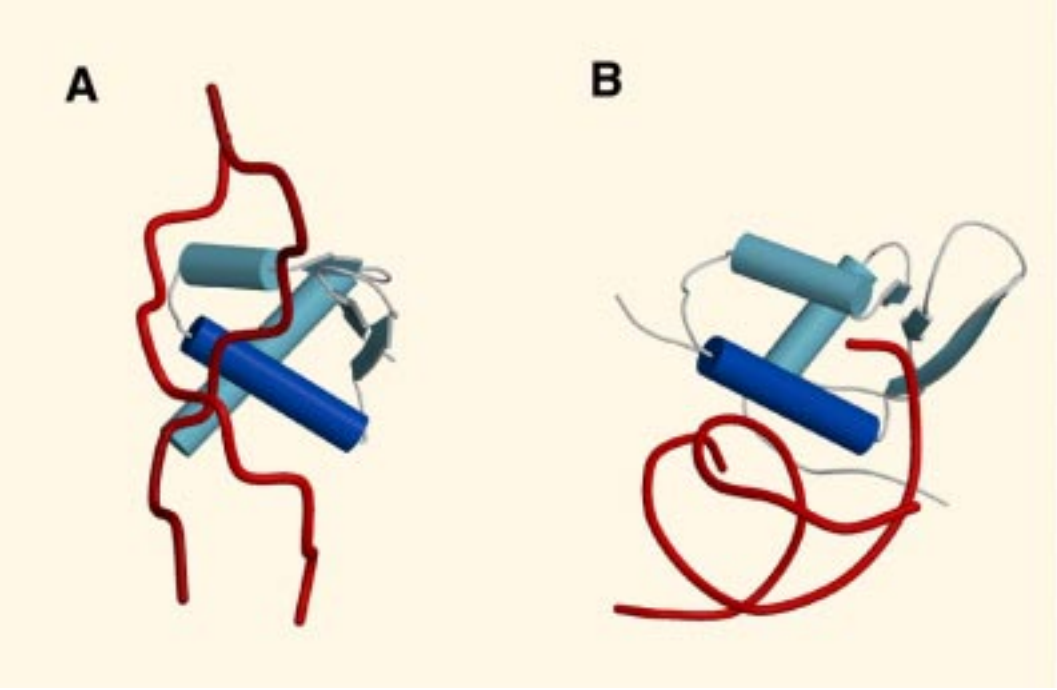

Figure 8.2.2 Docking modes of Zα and of a canonical B-DNA-binding HTH protein.
Schematic drawing of the Zα-DNA complex (**A**) in comparison to the HNF-3γ-DNA interaction (**B**) (Clark *et al.*, 1993). The proteins are oriented with their helix-turn-helix units in the same orientation. The recognition helix α3 of the HTH motif is shown in dark blue; the backbone trace of the DNA is red. The DNA axes are radically different due to the different 'angle of attack' the two proteins use to bind DNA.

In Zα, recognition helix α3 is positioned on the outer surface of the DNA with residues of the first two helical turns contacting the DNA. These residues are in the same positions as α3 residues used for B-DNA binding by CAP, HNF-3γ, DtxR, DP and E2F (Fig. 8.2.1). In contrast to B-DNA-binding HTH proteins with α/β topology, the first helix of the HTH motif and the preceding helix in Zα are not involved in DNA contacts. Instead, the C-terminal β-hairpin

contributes a second set of interactions. This β-hairpin is structurally well-defined, and is the region that deviates to the greatest extent from other HTH proteins with α/β topology. The cluster of highly conserved residues in the hairpin is specific to the Zα family (Fig. 8.1.1) and necessary for Z-DNA binding (Schade *et al.*, 1999). Unlike most structurally defined B-DNA-binding HTH proteins, the interaction seen with Zα is conformation- rather than sequence-specific. The base contact to carbon 8 of G4 (Fig. 7.8.4) specifies the *syn* as opposed to the *anti* conformation of this base, but any base in *syn* conformation at that position would expose a similar surface. These findings provide a structural basis for biochemical results, which describe a conformation–specific binding behavior for the Zα domain (Herbert *et al.*, 1998).

Comparing Z-DNA binding by Zα with B-DNA binding by HTH proteins of the same class gives the impression that Zα is the big exception to the rule. Several points must be stressed to correct this impression. First, B-DNA has been used above very generally to describe right-handed DNA. However, the variations among the B-DNA ligands are quite dramatic. Most of the complex structures show that the DNA ligand are bent to various extents — up to 90 degrees in the case of the CAP-DNA structure (Schultz *et al.*, 1991). Thus, the common feature of B-DNA binding is in fact limited to the observation that the recognition helix penetrates into the major groove. Additional contacts vary, the angle between the recognition helix and the DNA helical axis varies and there is also no simple code of recognition (Matthews, 1988), i.e., a direct correlation between a residue and a specific base pair contact made. Therefore it is extremely difficult to predict the conformation and sequence of a potential binding site, if, for example, the structure of a HTH protein without substrate is known. A second consideration about binding specificities of HTH proteins results from recently published experiments, where the HTH motif occurred in unexpected contexts. In the endonuclease Fok I, two HTH domains, D1 and D2, are used to bind its specific restriction site; interestingly, the recognition helix in D2 contacts the DNA in a non-canonical way, using a very different docking mode (Wah *et al.*, 1997). In the case of the ribosomal protein L11, which is structurally similar to the homeodomain class of HTH proteins, the ligand is RNA instead of DNA (Conn *et al.*, 1999; Wimberly *et al.*, 1999). For the homeodomain protein bicoid, sequence-specific binding of both DNA and RNA

is necessary for biological function (Chan & Struhl, 1997). These examples show how small variations in the binding site of HTH proteins can result in very different ligand specificities. This broad use of the HTH motif means that biochemical data must be combined with structural information to establish the specificity of binding of a given protein. The structure of the Zα-Z-DNA complex has features that explain the high affinity for Z-DNA. On the other hand, there are no steric or obvious electrostatic conflicts, which exclude the possibility that other ligands might be bound alternatively.

8.3. Comparison of the Proteolytic Experiments with the Crystal Structure of Zα

At the beginning of this study, a Zα domain, functional in binding Z-DNA in gel retardation assays, was available. This domain of residues 121-197 of ADAR1 was conformationally inhomogeneous as concluded from ion exchange chromatography. Therefore, limited proteolysis was chosen to gain information about the structural organization of the Z-DNA binding domain in ADAR1. A minimal stable and functional domain Zα was mapped to residues 133-204 of ADAR1. The difference in stability was thought to arise from additional protease-resistant residues at the C-terminus. The crystal structure cannot explain this hypothesis, since the residues C-terminal to I^{197} appear to be mostly disordered. It is possible that these disordered residues are displaced, due to crystal packing forces, from their position on the domain surface in solution. This would explain their contribution to the stabilization of the domain *in vitro*. The truncated N-terminus, defined by limited proteolysis, makes sense. The compact domain is a structural entity, so additional secondary structures (residues 122-130 were predicted to form a helix (Herbert *et al.*, 1997)) appear to be unnecessary for autonomous folding.

Another interesting aspect of the limited proteolytic experiments is the behavior of the Zβ domain. The alignment of the sequence of Zβ with that of Zα (Fig. 8.1.1) gives a very likely prediction of the Zβ fold. It is presumably very similar to Zα, since both sequences have a high degree of conservation among residues forming the hydrophobic core in Zα. It appears that α1 is shorter in Zβ, and that the loop region connecting α1 and α2 is one residue longer. In addition, the beginning of α3 is slightly different, due to the insertion of an

alanine residue in Zβ. Since there are only these slight modifications in the overall fold, it is intriguing that Zβ is so much more susceptible to protease digestions. A plausible explanation is that one of these slight changes creates a flexible hinge in a region that is rather well ordered in Zα. If such a hinge creates a protease sensitive site, this could explain the instability of Zβ. The compact domain would lose its stability immediately when cleaved in a single position at any place within the core fold, since every part of the domain contributes to the entire three-dimensional structure.

8.4. The *in vivo* Function of Zα

This crystal structure reveals the structural basis for the high-affinity binding of Z-DNA observed for Zα *in vitro.* Zα is part of ADAR1, an RNA editing enzyme. RNA editing is still only partially understood, which also affects understanding of the possible role of what Z-DNA binding might do in this process. A plausible model for the Z-DNA binding activity in ADAR1 has been proposed (Fig. 2.2.2). However, this model's validity has not yet been proved. There are various indications that the original hypothesis for the Z-DNA binding activity, proposed by Herbert & Rich (1996), might demand modifications. ADAR1 occurs in various splice variants in the cell, including variations in the N-terminus. In fact, the majority of the cellular protein lacks Zα and starts at an alternate start codon at the N-terminus of Zβ. The full-length 1221 aa protein is synthesized in an interferon-dependent manner (George & Samuel, 1999), signaling a connection with antiviral response mechanisms of the cell. According to the findings presented in this thesis, only the full-length protein can tightly bind Z-DNA, when the entire Zab domain is present. The truncated form of ADAR1 probably cannot bind Z-DNA tightly. Another indication that Z-DNA binding might be relevant in antiviral response is that the only close homologue to Zα and Zβ found in the database is the gene product E3L from the poxvirus *Vaccinia* (Herbert *et al.*, 1997). This protein consists of two domains, a C-terminal dsRNA binding domain and an N-terminal Zα-related domain (Chang & Jacobs, 1993). E3L is a dimeric protein in solution and binds Z-DNA with apparent nanomolar affinity (unpublished data). No function has been established for the Z-DNA binding activity in E3L so far. Yet it is possible that Z-DNA formation occurs in the viral life cycle in a

manner unknown until now. A model could emerge, once this field of research has been investigated further. Without *in vivo* data available, it is of course not exclusive that Zα or Zab might have a non Z-DNA related function too. However, the high degree of conservation among residues involved in Z-DNA binding in the Zα family makes it very unlikely that this possibility would emerge.