2. Scientific Background

2.1. Z-DNA

After the discovery of Z-DNA in the crystal structure in 1979 (Wang *et al.*), it was of central interest first to understand Z-DNA formation *in vitro* and *in vivo*. Elucidation of these principles would allow for addressing the key questions: whether Z-DNA indeed occurs *in vivo* and, most importantly, whether it exhibits a function.

The Z-form describes a left-handed double-helical DNA conformation with central Watson-Crick-base pairing between the anti-parallel running strands. The Z-helix has a smaller diameter than its B-DNA counterpart and a full helical turn comprises 12 bases, as opposed to 10 in B-DNA. Z-DNA can be formed reversibly from B-DNA sequences. For this transition to occur, every second base, alternating in both strands, must rotate by 180 degrees around the base–sugar linkage (Fig. 2.1.1). This process transforms nucleosides from the anti to the syn conformation. Whereas in B-DNA all nucleosides are in syn, in Z-DNA syn and anti conformation alter. The syn conformation is sterically unfavorable for pyrimidine bases; therefore Z-DNA formation usually is accompanied by a purine pyrimidine alternation within the base sequence, although there are exceptions to this rule (Wang, *et al.*, 1985; Schroth *et al.*, 1993). Bases in syn nucleosides are presented on the helical surface in a way distinct from bases in anti nucleosides. Carbon C8 of syn purine becomes exposed in Z-DNA, as opposed to B-DNA, where it is buried (Fig. 2.1.2).

A second characteristic feature of Z-DNA, besides the anti/syn alternation, is the sugar puckering. It is C3'-endo for the syn nucleosides, in contrast to the anti nucleosides, which are puckered C2'-endo, as is B-DNA. The consequence of the anti to syn transition is that the interphosphate distance on the DNA is shortened from 7.0 Å to 5.9 Å in every second position. Another consequence is that, strictly speaking, the Z-helix has a dinucleotide repeat rather than a mononucleotide repeat as present in all other known DNA forms. This explains the 'zig-zag' course of the backbone.

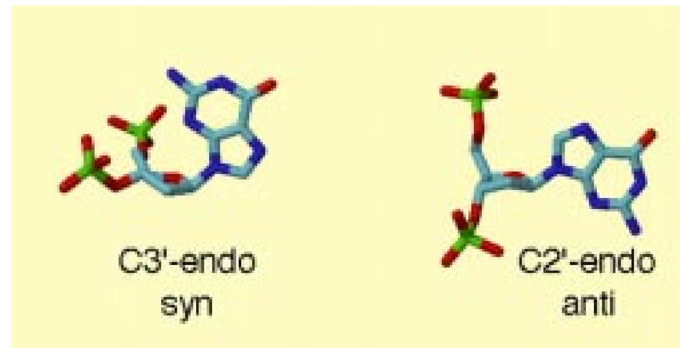

Figure 2.1.1 Nucleotide conformations.
In Z-DNA (left), every second nucleotide is in the syn conformation with C3'-endo sugar pucker. In B-DNA (right), nucleotides are in the anti conformation with C2'-endo sugar puckers.

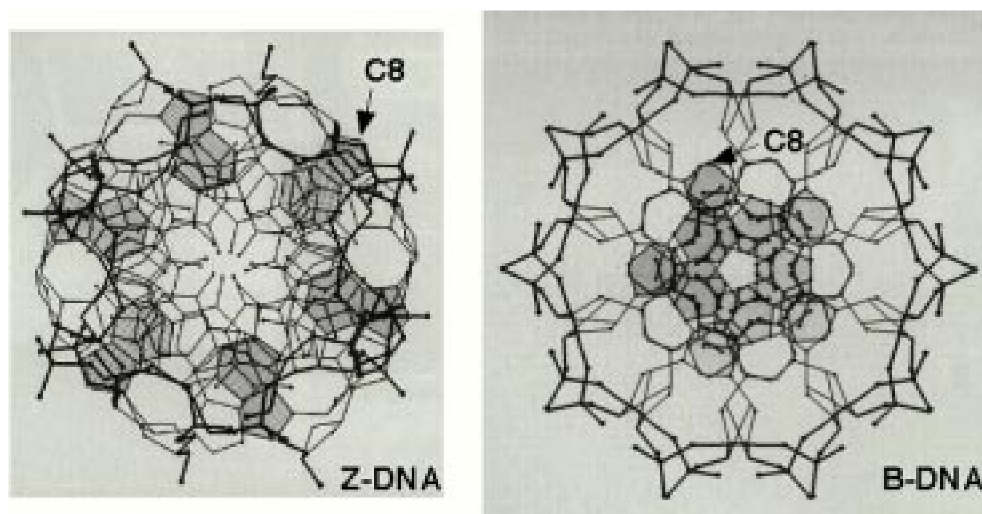

Figure 2.1.2 Exposed carbon C8 position in Z-DNA.
View down the helical axes of Z-DNA and B-DNA. In Z-DNA, the carbon C8 atom of syn-purines is exposed on the helical surface, as opposed to B-DNA, where it is inaccessibly buried.

Under usual physiological conditions, Z-DNA is energetically less favorable than B-DNA. Consequently, the equilibrium between B- and Z-DNA is almost entirely shifted towards B-DNA. Yet, various *in vitro* conditions can be applied, which shift the equilibrium towards Z-DNA (i.e. divalent cations, high salt, reduced solvent polarity, chemical bromination; Rich *et al.*, 1984). In the context of a living cell, these conditions are largely irrelevant. It is the dynamic nature of DNA that produces conditions stabilizing Z-DNA *in vivo.* It has been shown that negative superhelical stress in DNA induces Z-DNA formation, thereby relaxing the DNA (Peck *et al.*, 1982). Gene transcription in metabolically active

cells puts the DNA under constant torsional stress. A transcribing RNA polymerase induces positive and negative supercoiling in DNA (Liu & Wang, 1987). The transient formation of negative supercoiling in the vein of the moving RNA polymerase could possibly be relaxed by Z-DNA formation. For the *c-myc* gene it has been shown that Z-DNA formation occurs in a transcription-dependent way *in vivo* (Wittig *et al.*, 1992). It has to be mentioned however that the mechanics of the B-Z transition are not yet understood. It is a sterically rather complicated process, since base pairs have to rotate by 180 degrees.

Whether Z-DNA formation simply occurs without consequence or whether it also exhibits a function in the cell is an important question. Since B- and Z-DNA are structurally so different, it is conceivable that the cell would use the Z-signal in some way. To exhibit a function, the signal produced by Z-DNA formation must be transduced. This task could most likely be fulfilled by a protein. Hence, large efforts were spent on finding candidate proteins. Due to practical reasons, it is difficult to identify such Z-DNA specific proteins. In the *in vitro* binding assay, the DNA needs to be stabilized in the Z-conformation in a way that best mimics the situation *in vivo.* After numerous unsuccessful attempts by various laboratories, a promising binding assay was developed (Herbert & Rich, 1993): a short radiolabeled DNA oligomer, held in the Z-conformation by chemical bromination, was used for electrophoretic mobility shift assays to search for specifically binding proteins from eukaryotic cell extracts. An activity that bound with high affinity and specificity was found (Herbert *et al.*, 1993). It was isolated and revealed to be an N-terminal fragment of the editing enzyme, double-stranded RNA (dsRNA) dependent adenosine deaminase ADAR1 (Herbert *et al.*, 1995). This fragment, Zα, bound with nanomolar affinity to Z-DNA and the binding was not competed by a 10000-fold mass excess of unspecific B-DNA or RNA (Herbert *et al.*, 1997).

2.2. Double-Stranded RNA Adenosine Deaminase ADAR1

ADAR1 is a member of a family of dsRNA dependent adenosine deaminases (Maas *et al.*, 1997). These enzymes act on dsRNA substrates, where they deaminate adenines to produce inosines. This reaction can take place in mRNA,

where the translation machinery reads the edited inosine not as an adenine, but instead as a guanine. The result is that the editing event introduces a codon change in the transcript. In consequence, protein isoforms can be generated from a single gene. The occurrence of this form of RNA editing, the site-selective modification of mRNA, has been demonstrated for a number of genes in higher eukaryotes. ADARs are involved in producing functionally important isoforms of mammalian serotonin receptors (Burns *et al.*, 1997), several mammalian glutamate receptors (Sommer *et al.*, 1991, Lomeli *et al.*, 1994) and the virally encoded hepatitis delta antigen (Polson *et al.*, 1996). A to G transitions also have been observed within codons of several other viral and cellular transcripts, although the function of these has not yet been determined (Bass, 1997).

It is of interest that in the case of the glutamic acid receptor subunits, modification of the exon is guided by folding of a downstream intron back onto the exon to form the dsRNA-editing substrate (Brusa *et al.*, 1995). Therefore, editing must occur early after transcription in the nucleus, before introns are removed.

ADAR1 is a 136kD protein comprising multiple domains. Similar to the other members of the ADAR family (Maas *et al.*, 1997), ADAR1 has a catalytic domain and multiple dsRNA binding domains (Fig. 2.2.1).

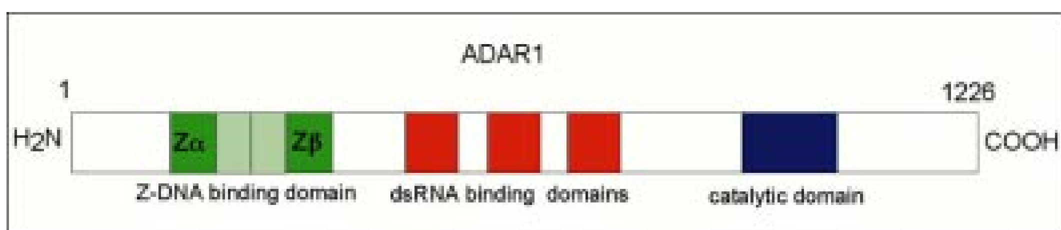

Figure 2.2.1 Domain organization of the dsRNA specific adenosine deaminase ADAR1.

Characteristic and distinctive for ADAR1 is the extended N-terminus harboring the Z-DNA binding domain, not present in the related deaminases ADAR2 (Melcher *et al.*, 1996a) and ADAR3 (Melcher *et al.*, 1996b). The area of interest is composed of two largely similar sequences, Zα and Zβ, which are separated by a linker region. What is the functional connection between Z-DNA binding and RNA editing? A plausible model has been proposed (Herbert & Rich, 1996); as noted above, pre-mRNA editing must take place early after transcription

occurs. The question arises: how is ADAR1 targeted to those early sites of transcription? The hypothesis is that Z-DNA formation within appropriate sequences, in the vein of a moving RNA polymerase, is used as a binding site for ADAR1 in order to bring the enzyme close to its place of action (Fig. 2.2.2).

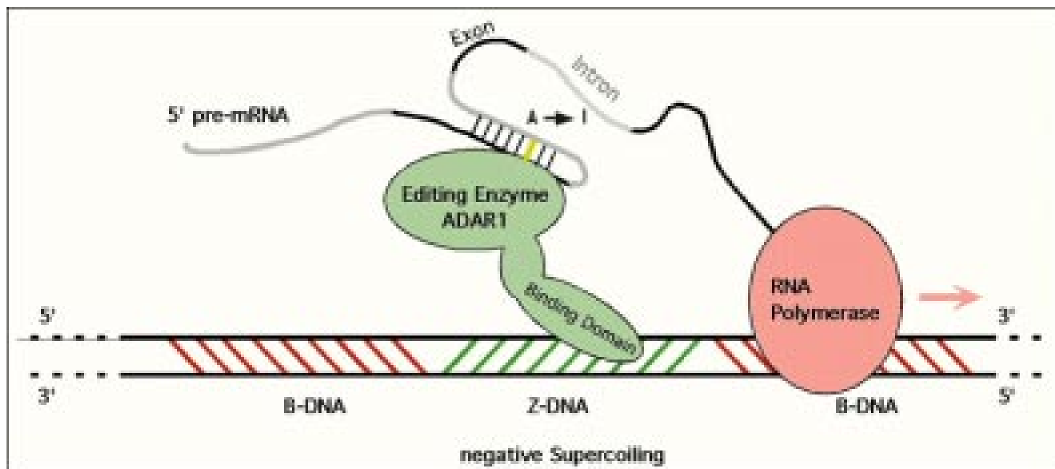

Figure 2.2.2 Functional model for Z-DNA binding by the RNA editing enzyme ADAR1.
RNA editing sites often are formed in pre-mRNA between an exon and an intron folded back on it. Deamination of a specific adenine to inosine occurs, which is read by the translation machinery as guanine. The result is a codon change, which leads to a site-specifically mutated protein. Since editing has to take place before splicing occurs, Z-DNA, which can be formed transiently behind a moving polymerase, might target ADAR1 to its place of action (Herbert and Rich, 1996).

This model is attractive, but there might be other models emerging. Recent studies have shown that ADAR1 exists in various splicing variants (Liu *et al.*, 1997). In addition two promoters regulate the transcription of a short form of ADAR1, starting just at the N-terminus of Zβ, or the full-length form, which is generated interferon-dependently (George & Samuel, 1999). The interferon dependency suggests that full-length ADAR1 might be involved in a not-yet understood anti-viral defense mechanism. Unselective adenosine deamination (hypermutation) of viral dsRNA in the cytosol has been observed in various cases (Cattaneo, 1994; Kumar & Carmichael, 1997). This form of RNA editing could be part of a defense strategy of the infected cell. Currently, different scenarios for the function of ADAR1 *in vivo* are under investigation.

2.3. Limited Proteolysis

For crystallographic purposes it is most desirable to have a biological macromolecule that is uniform and rather inflexible. Protein domains are usually well-structured regions of 50 to 200 amino acids (Schulz & Schirmer, 1979). Larger proteins are built from multiple, mostly independently folded, domains. The regions connecting those domains are often flexible and solvent exposed. Limited proteolysis is a classical approach to define domain organization (Porter, 1999; Jovin *et al.*, 1977; Roy *et al.*, 1996; Nakagawa *et al.*, 1997). It takes advantage of the fact that site-specific proteases will cleave proteins preferentially in solvent exposed unstructured regions, rather than within a folded domain. Thus, treating a large, multi-domain protein with limited amounts of site-specific proteases will generate proteolytic fragments representing the stable subdomains. In crystallography, this technique has been applied with great success to grow x-ray-quality crystals of the protein of interest.

2.4. Protein-DNA Binding Assays in Solution

2.4.1. Electrophoretic Mobility Shift Assay

The electrophoretic mobility shift assay (EMSA) is a powerful method to study protein-DNA interactions in solution (Lane *et al.*, 1992). It is based on the fact that the electrophoretic mobility of DNA oligomers through a polyacrylamide gel is altered when a protein is bound. A mixture of the protein-DNA complex in equilibrium with the free DNA and free protein is loaded onto a running gel. The species are then separated according to their electrophoretic mobility, which is a function of size, shape and charge. A complex mixture of bound and unbound DNA can therefore be separated by this method; determination of the protein binding affinity is possible if the DNA substrate is radioactively labeled.

2.4.2. Circular Dichroic Spectroscopy

In circular dichroic (CD) spectroscopy the sample is excited with circularly polarized, monochromatic light. The light wave can be regarded as a

combination of two linearly polarized waves, 90 degrees phase shifted and mutually perpendicular. In chiral molecules, the two light waves are absorbed to a different extent. As a result, the combined wave oscillates with a changed amplitude in a direction different from the original. CD spectroscopy is an excellent low-resolution technique to study conformational changes in biological macromolecules. A CD experiment gave the first picture of the unusual conformation of a $(dCdG)_n$ polymer at high ionic strength, later called Z-DNA (Fig. 2.4.1.1)

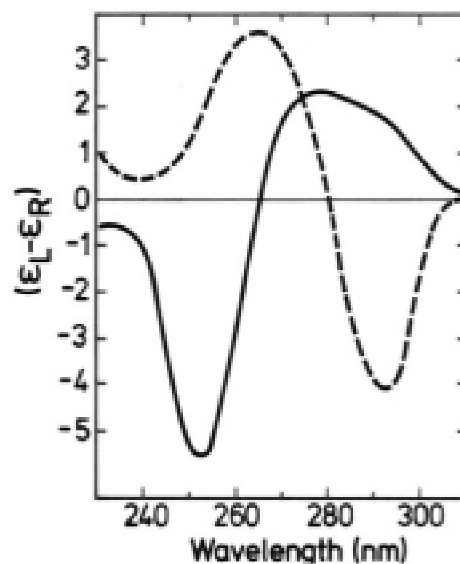

Figure 2.4.1.1 Circular dichroism spectra of poly(dG-dC) at 0.2 M NaCl, pH 7, 25 °C (solid line) and after addition of solid NaCl (dashed line). From Pohl & Jovin, 1972.

2.5. Principles of Macromolecular X-ray Crystallography

2.5.1. Scattering of X-rays

Depending on the number and positions of the electrons about the nucleus, the atoms in a molecule deflect an X-ray beam. The scattering effect can be expressed by the atomic scattering factor $f(\mathbf{S})$, where $\rho(\mathbf{r})$ describes the electron density at position r and S is the difference vector between incoming and scattered x-rays.

$$f(\mathbf{S}) = \int \rho(\mathbf{r}) \exp(2\pi \mathbf{r} \cdot \mathbf{S}) d\mathbf{r}$$

The total scattering from a unit cell in a crystal is the sum of the scattering contribution of the individual atoms and is expressed by the structure factor **F(S)** summed over all atoms I in the unit cell.

$$\mathbf{F(S)} = \sum_i f_i \exp(2\pi i \mathbf{r}_i \cdot \mathbf{S})$$

To determine the scattering by the crystal as a whole, the scattering of all the individual unit cells with respect to a single origin must be added. For a unit cell with its origin at position $\mathbf{r} = t\mathbf{a} + u\mathbf{b} + v\mathbf{c}$, in which t, u and v are integers, the scattering is defined as

$$\mathbf{F(S)} \times \exp(2\pi i \mathbf{r} \cdot \mathbf{S})$$

The total wave **K(S)** scattered by the crystal is therefore obtained by the summation

$$\mathbf{K(S)} = \mathbf{F(S)} \times \sum_i \exp(2\pi i t \mathbf{a} \cdot \mathbf{S}) \times \sum_i \exp(2\pi i u \mathbf{b} \cdot \mathbf{S}) \times \sum_i \exp(2\pi i v \mathbf{c} \cdot \mathbf{S})$$

Due to the large number of unit cells in a crystal, the result of each summation is almost always zero unless **a·S, b·S** and **c·S** are integer values, and these solutions are known as the Laue condition. The integer values are defined as h, k and l, and represent the Miller indices of the reflections from the crystal. In 1913, Bragg demonstrated that the scattered or diffracted beams could be regarded as reflections from an imaginary series of planes in the crystal. From this interpretation, the Bragg Law emerges:

$$2d \sin \Theta = n\lambda$$

where λ represents the wavelength of the incoming x-ray, Θ is the reflection angle at the lattice plane and d denotes the distance between lattice planes. The Miller indices, hkl, define the so-called reciprocal lattice. Each lattice point in the reciprocal lattice corresponds to a plane in the real lattice with the indices hkl. The length of a vector in a reciprocal lattice is $1/d_{hkl}$ where d_{hkl} is the distance between parallel planes in the crystal lattice.

2.5.2. Calculation of the Electron Density

Instead of summing over all separate atoms, integration over all electrons in the unit cell can be carried out, which yields an expression for the structure factor of

$$\mathbf{F(S)} = \int_{cell} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r} \cdot \mathbf{S}) d\mathbf{r}$$

where $\rho(\mathbf{r})$ is the electron density at position **r** in the unit cell. Fourier transformation of this expression leads to

$$\rho(\mathbf{r}) = \int_{cell} \mathbf{F(S)} \exp(-2\pi i \mathbf{r} \cdot \mathbf{S}) d\mathbf{S}$$

Alternatively, by using the fractional coordinates x, y and z of the unit cell, $\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$, the structure factor amplitude $|F(hkl)|$ and the phase angle of reflection *(hkl)*, $\alpha(hkl)$,

$$\rho(xyz) = \frac{1}{V_{cell}} \sum_{hkl} |F(hkl)| \exp(-2\pi i(hx + ky + lz) + i\alpha(hkl))$$

The electron density in the unit cell can therefore be calculated from the contributions of the structure factor amplitudes $|F(hkl)|$ and the phases $\alpha(hkl)$. The intensity of the diffracted beam is proportional to the square of its amplitude

$$I(hkl) \propto a|F(hkl)|^2$$

with the *a* term representing a constant that corrects for parameters such as absorption, data acquisition geometry, dependence of the diffraction on wavelength, etc. The phases α, however, cannot be obtained directly from the diffraction pattern and have to be deduced indirectly. This difficult task is often referred to as the phase problem in crystallography.

2.5.3. Solution of the Phase Problem

Four prevalent methods have been developed for the solution of the phase problem:

Direct methods are mainly used in small molecule crystallography. The derivation of phases is based on statistical probability theory. These methods are still in an early stage of development for application to macromolecules.

Molecular replacement is the most expeditious procedure for structure determination, but a structure with a very similar skeletal motif must be available. Initial phasing is based on phase approximation taken from the data of the comparable structure.

Multiple wavelength anomalous diffraction (MAD) can be used when the structure contains atoms that exhibit strong anomalous scattering. In these atoms, the electrons cannot be regarded as free electrons, and the scattering of the beam by these atoms will be dependent on the wavelength of the x-ray. The phases for the structure can be determined from a comparison of the diffraction data collected at several different wavelengths.

Multiple isomorphous replacement (MIR) requires the incorporation of heavy atoms (atoms with high atomic number, i.e. Au, Hg, I, Pt) into the unit cell either through co-crystallization, soaking or by covalent attachment to the macromolecules. For this method to be successful, at least two different derivatives must be crystallized in the identical space group with identical unit cell dimensions. The intensity differences between the native and the derivative diffraction patterns are thus exclusively due to the attached heavy atoms. Determination of the heavy atom positions provides a starting point for derivation of the macromolecule phase angles. Presently, this technique is the most commonly used method for the determination of a novel structure.

For the determination of the complex structure of the Zα domain bound to left-handed Z-DNA described herein, the single isomorphous replacement technique, including anomalous scattering (SIRAS), was used. This is in a way a special application of the MIR technique, where the intrinsic ambiguity of the phases obtained from a single derivative is solved by the complementary information from the anomalous signal of the heavy atom.