1. Introduction

The genetic information of all prokaryotic and eukaryotic organisms is stored in the nucleic acids. Structural polymorphism of nucleic acids is a biologically significant property and reflects biological function. Equally important are proteins that can be regarded as a precision molecular machines in itself. These machines operate in very systematic manner in the living cell. Proteins control many processes inside the living cell utilizing a multitude of interactions of chemical and physical nature and structural flexibility. Any understanding of the biological functions of nucleic acids and proteins requires an extended knowledge of their properties. Knowledge of the fundamentals enables to develop new medical therapies and synthesize new chemical compounds that could be applied in the treatment of various diseases.

Very powerful techniques of structural biology are x-ray crystallography and nucleic magnetic resonance (NMR) which provide the three dimensional structures of biological macromolecules at atomic resolution. However, the complex nature of proteins and DNA demands for additional techniques that can provide more information about all the properties of interest. For example, the stability, the relevance of the structure to environmental conditions, and the kinetics of reactions request a multitude of further methods for their determination. Raman spectroscopy is one of the techniques which can be applied to extend the information obtained by NMR and x-ray crystallography. It is physically based on the fact that all atoms of a molecule vibrate. Vibrations depend on the mass, the charge, the binding strength, and any intra- and intermolecular interactions of the atoms and groups of atoms of a molecule. Particular vibrations are influenced by the force field formed by the surrounding, and reflect molecular conformation as well. In complexes, the approach of one molecule to another obviously perturbs vibrations in and near the contact area.

By means of Raman difference spectroscopy, one can study the changes of the conformation before and after the approach of molecules and thereby monitor the adaptability and flexibility of the studied system. In this work proteins and DNAs were investigated with special emphasis on changes of their conformations caused by the interactions upon complex formation. Four protein–DNA complexes (ω_2–operator DNA, Arc–operator DNA, KorB–operator DNA, and Sac7d–decamer d(GAGGCGCCTC)$_2$) are characterized by Raman spectroscopy.

1.1 Bacterial plasmids

Plasmids are extrachromosomal, autonomously replicating genetic elements that are profoundly important for bacterial adaptability and persistence. They exist in circular or linear form in the horizontal gene pool of bacteria. Plasmids harbor genes that encode proteins involved in a variety of biological functions and provide functions that might not be encoded by chromosome (*1*). Although plasmids can be highly beneficial to host bacteria allowing them to survive in hostile environment, killing mechanisms are also carried by plasmids.

1.1.1. Plasmid Maintenance

Naturally occurring plasmids are usually genetically stable. Their stability is often ensured in spite of very low copy numbers due to the presence of replication-control mechanisms that actively prevent loss of the plasmids and direct them to new daughter cells at cell division. Gene cassettes that prevent plasmid loss can be divided into three classes (*2*): (i) centromere-like systems that actively secure ordered segregation of replicons prior to cell division (active partition process) (*3*), (ii) site-specific recombination systems that actively resolve tandem plasmid multimers into monomers (*4*), and (iii) cassettes that mediate killing of newborn, plasmid-free cells resulting from failure of the first two systems to secure plasmid maintenance. These plasmid-directed events resulting in selective killing of cells that have failed to inherit plasmid copy were identified in 1980s and termed postsegregational killing (PSK) (*5*).

1.1.2. Plasmid Partitioning

To ensure precise distribution of plasmid copies to prospective daughter cells, an active partitioning (Par) is employed during the process of cell division for plasmids having moderate or low-copy numbers. In contrast, high copy number plasmids rely on random partitioning. If the plasmids are distributed randomly, it is the number of plasmid copies n inside the cell which determines the probability P_0 that one of the daughter cells will be plasmid free: $P_0 = 2^{(1-n)}$ (*6*). Therefore, a cell having 31 plasmid copies at the time of division would produce one plasmid-free cell in 10^9 cells, whereas for 4 plasmid copies, the chance for the cell to be plasmid-free is 1 to 8.

Three components, partition proteins ParA, ParB and *par* site (*parS*) are required for partitioning mechanism of moderate or low-copy number plasmids. ParB protein binds to

cis-acting sequences (*parS*) known as centromere-like region (*7*). ParA is characteristic for showing ATPase activity and providing the energy necessary for the transport process (*8*). Chromosomal homologues of ParA and ParB proteins exist in many bacteria (*9*).

1.1.3. Postsegregational Killing Mechanism

PSK systems eliminate plasmid-free cells that might otherwise outgrowth those carrying the plasmid burden. Two types of PSK have been described in detail at molecular level (*10-11*). In both cases, these toxin-antitoxin (TA) mechanisms also known as "plasmid addiction" systems comprise a stable proteic toxin coupled with metabolically unstable antitoxins (antidotes). The relative instability of the antitoxin results in the activation of the stable toxin in progeny in which the toxin- and antitoxin-encoding plasmid has disappeared because of replication error or other defect of plasmid maintenance.

One type of mechanism is composed of unstable antisense-RNA species preventing the translation of protein-encoding mRNAs. In plasmid free cells, the translation machinery produces toxic protein from the mRNA transcript leading to cell death (type I). The analysis of the *hok-sok* locus of plasmid R1 has not only provided important insight into PSK mechanism but also into RNA-RNA interaction (*10*). The absence of the *sok* antitoxin RNA in plasmid-free cells results in the accumulation of the *hok* mRNA that is translated into the 52-amino acid, toxic Hok protein. Hok targets the cell membrane in similar way to holin proteins produced by some bacteriophages before cell lysis (*10*).

In the second type of PSK mechanism, the regulators are cognate toxin and antitoxin proteins. The higher population of less stable antitoxin binds toxin that is than unable to act within the cell. If a plasmid free cell arises, proteases degrade the antitoxin and the toxin becomes available to target an essential intracellular host factor to eliminate the cell (type II).

In general, TA loci are organized into operons in which the first cistron encodes the antitoxin and the second the toxin (Figure 1.1). Usually, the antitoxins are smaller than toxins whereas the reverse case is very exceptional (Table 1.1) (*2, 11*). Very often the antitoxins or TA complexes participate in autoregulation of the TA operon. One exception in TA loci is the ω-ε-ζ operon, in which ε encodes an antitoxin while ζ encodes a toxin. None of the proteins encoded by ε and ζ genes are involved in transcriptional regulation. Instead of that, ω gene indirectly controls the expression of ε and ζ genes (*12-13*).

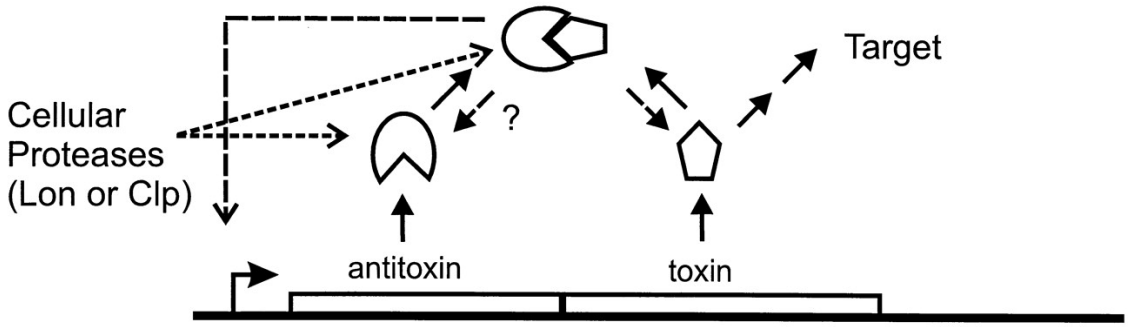

Figure 1.1: General genetic and functional setup of the TA loci. The antitoxin deactivates the toxin by forming tight complexes. The TA complex binds to the operator in the promoter region and regulates its transcription. If the cellular proteases degrade the antitoxin, then the free toxin attacks its target (Table 1.1). The question mark means that it is yet unknown whether the proteases degrade free antitoxin or complexed with toxin. (Adopted from ref 2)

Table 1.1: Major type II TA loci identified on plasmids[a]

Plasmid	Bacterium	Toxin (aa)	Antitoxin (aa)	Toxin target
F	E. coli	CcdB (101)	CcdA (72)	DNA gyrase
P1	E. coli	Doc (126)	Phd (73)	Translation?
RK2	E. coli	ParE (103)	ParD (83)	DNA gyrase
R1	E. coli	Kid (110)	Kis (84)	DnaB
pTF-FC2	A. ferroxidans	PasB(90)	PasA (74)	ND
pSM19035	S. pyogenes	ζ (287)	ε (90)	ND
Rts1	E. coli	HigB (92)	HigA (104)	ND
P307	E. coli	RelE (95)	RelB (83)	m-RNA[b]
pMYSH6000	S. flexneri	MvpT (133)	MvpA (75)	ND
pRUM	E. faecium	Txe (85)	Axe (89)	ND

[a] Prototypical TA loci are only listed; many other plasmids and chromosomes produce homologs of these proteins. aa, amino acids; ND, not determined. (Adopted from ref 11)
[b] Target of RelE has been recently found (*14*).

1.1.4. *inc18* Family and Plasmid pSM19035

The low copy number and broad host range plasmid of the *inc18* family (pIP501, pAMβ1, and pSM19035) from Gram-positive bacteria share a large sequence identity (>90%) in their replication (Cop, RNA III, REP, ori) and stability (SegA and SegB) regions (*15-16*). Replication of plasmid is controlled by two negative regulators, each of which reduces the amount of mRNA coding for Rep, the initiation protein. First one, Cop protein, limits the amounts of *rep* mRNA either directly as a repressor or indirectly, by affecting the transcription of stable antisense RNA (RNA III) (*17*). RNA III induces transcriptional attenuation within the leader region of *rep* mRNA (*16-17*). SegA and SegB play a role in the segregation stability of pSM19035 (*18*). SegA assures that oligomeric forms are converted into monomers. The SegB region is required for better-than-random

plasmid pSM19035 distribution, including *δ*, *ω*, *ε*, and *ζ* genes (*12, 18*). The δ protein is a homolog with ATPases involved in partitioning of diverse bacterial plasmids and bacterial chromosomes (*18-19*), whereas *ε* and *ζ* gene products belong to PSK type II (*20*).

1.2. ω Protein

Streptococcus pyogenes pSM19035-encoded ω protein represses not only its own synthesis, but also blocks transcription from *copS* and *δ* genes, thereby controlling plasmid maintenance (*12*). The ω-mediated decreased synthesis of CopS will increase transcription of *rep* mRNA and indirectly decreases transcription of antisense RNA III (*12*). The *rep* mRNA will be further translated into Rep protein that initializes the plasmid replication. An excess of ParA-like protein δ actively contributes to the pSM19035 partition (*12, 17, 21*). The *ω*, *ε*, and *ζ* genes form an operon with two promoters, Pω reading all three genes and Pε only *ε* and *ζ* genes. Thus, ω protein indirectly controls expression of the *ε* and *ζ* genes of ω-ε-ζ operon which products prevent the appearance of plasmid-free segregants. Altogether, ω protein is a global regulator actively engaged in plasmid copy number control and stable maintenance (*12*).

1.2.1. Properties and Crystal Structure

The purified ω protein is composed of 71 amino acid residues, with mainly α-helical (42%) folding, and occurs as homodimer in solution (ω_2, molecular mass 16 kDa) with a K_d of 3.2 μM (*22*). Spectroscopic studies have shown that ω_2 unfolds and refolds reversibly depending on urea concentration. CD spectroscopy indicated half transition temperatures of thermal unfolding, T_m, between ~43 and ~78°C depending on the ionic strength of the buffer (*22*).

The crystal structure of ω protein exhibits two ω monomers related by a non-crystallographic 2-fold rotation axis forming a homodimer (ω_2) that occupies the asymmetric unit (*21*). Each ω monomer is folded into two helices A and B, and one β-strand forming an antiparallel β-ribbon in the dimer (Figure 1.2). The 20 N-terminal amino acids were not identified probably because of loss by proteolysis during the crystallization (*21*). The dimer structure is stabilized by several monomer-monomer contacts between helices B and B' including hydrogen bonds, salt bridges and interactions among hydrophobic side chains (*21*). The ω_2 antiparallel β-ribbon is stabilised by normal main-

chain hydrogen bonds and additionally by hydrogen bonds and salt bridges among residues at both ends of the β-ribbon.

The ω protein belongs to the structural superfamily of MetJ/Arc repressors featuring a ribbon-helix-helix (RHH) DNA binding motif (*21*). Proteins of this family recognize the major groove of DNA with the β-ribbon of the RHH motif. Several structures of proteins of this family have been solved: Mnt (*23*), Arc (*24-26*), MetJ (*27*), and CopG (*28*). Structures of most of them have been also determined in complexes with their operator DNAs: Arc (*29*), MetJ (*30*), CopG (*28*), and NikR (*31*).

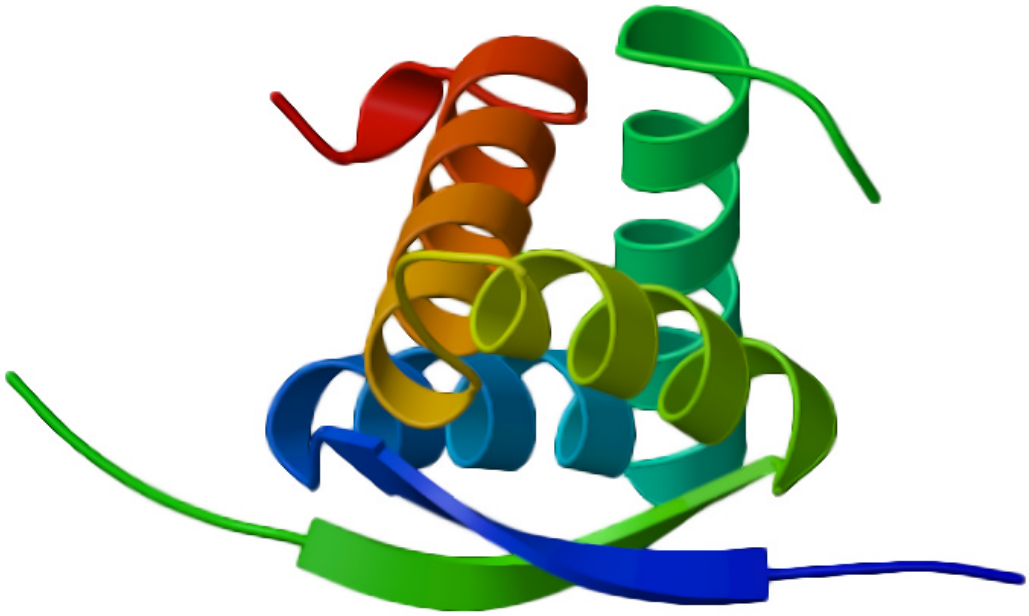

Figure 1.2: Ribbon diagram of the side view of ω protein homodimer. Subunit I and II are drawn in blue-green and green-red, respectively (*21*) (PDB 1irq).

1.2.2. Ribbon-Helix-Helix Binding Motif

The crystal structures of RHH or $(\beta\alpha\alpha)_2$ proteins in complexes with their operator DNAs reveal several similarities and also some differences. First, the operator sequences contain two or more binding sites that are usually arranged as inverted (Arc, CopG) or direct (MetJ) repeats. The proteins bind the DNAs as tetramers, or higher order oligomers, with dimers binding to each half-site of the operator and making cooperative interactions with the neighboring dimer. The dimer-dimer interactions are different depending on the spacing between the two β-ribbons in the protein–DNA complexes. The distance also determines the surface occluded by two dimers. Arc dimers interact with each other over a limited region (~325 $Å^2$/dimer) using the loop residues between helices A and B. The

dimers of MetJ pack against each other over a larger region (~550 $Å^2$/dimer) along the length of helix A (*32*). The CopG dimers interact over a region (~526 $Å^2$/dimer). Multiple contacts occur between the loop residues connecting all three major secondary structure elements and between almost entire helices B from both dimers (*28*).

Second, the right and left binding-site reveals general symmetry. Although the binding sites are symmetrical for Arc, CopG, and MetJ, only MetJ interacts symmetrically with its directly repeated cognate site. Both Arc and CopG interact asymmetrically with their palindromic targets.

Third, all the critical base contacts with the operator DNA are arranged by the amino acids of the antiparallel β-ribbon that lies flat in the major groove. Six amino acids from β-ribbon of Arc and MetJ dimers make hydrogen bonds with DNA bases on each binding site. In the CopG–operator DNA complex, only four hydrogen bonds are formed on each binding site.

Fourth, several of contacts are made between DNA phosphates and helix B of each protein subunit, which seems to be crucial for binding (*32*). The antiparallel β-ribbon lies in the major groove, but the fit is not perfect to prevent wobbling and twisting. To solve this problem, Arc, CopG, and MetJ make additional stabilizing hydrogen bonds with DNA phosphates (*28-30*). Additionally, both Arc and MetJ contain flexible N-terminal region. The N-terminal region of Arc is largely disordered in solution as well as in the crystal structure (*26, 33*). The N-terminal region of MetJ forms different conformations depending on the crystal environment (*27*). In both Arc and MetJ complexes, the first seven residues before the β-sheet adopt compact conformation and interact with DNA backbone. The N-terminal regions also partially mediate the DNA sequence specificity. A Mnt mutant that has the first 9 N-terminal residues of Arc instead of its own first six residues binds specifically to the *arc* operator (*34*).

Finally, operator–DNA structures of one tetramer binding to two half-sites of the operator show bending of ~50° for Arc and MetJ and 60° for CopG. In the CopG operator, the bend is produced by compression of both major and minor grooves facing the protein (*28*). In the *met* operator, the major grooves are compressed around the bound repressor antiparallel β-ribbons (*30*), whereas in the *arc* operator the major grooves are widened around the β-sheets and the minor groove is compressed near the center of the operator binding site (*29*). These bends are required for the cooperative interaction of bound dimers.

1.2.3. Binding Site and Properties of ω_2–Operator DNA Complex

The ω_2 protein binds specifically with high affinity to 7-bp repeats (5'-$^{A}/_{T}$ATCAC$^{A}/_{T}$-3', ⟶, top strand of the consensus motif) located in the upstream operator regions of *cop*S and δ genes and in that of the ω-ε-ζ operon (*12*). The 7-bp repeats (heptads) are arranged in different modes in *P*δ, *P*ω and *P*copS, e.g. in *P*ω in two copies of a block with two direct and one inverted 7-bp repeat, and one inverted 7-bp repeat downstream of the block symbolized by [(⟶)$_2$(⟵)]$_2$(⟵). The arrangements are [(⟶)$_2$(⟵)]$_3$ (⟶) in *P*cop and (⟶)$_8$(⟵)$_2$ in *P*δ. The sequences of upstream promoter regions of the *ω*, *copS*, and δ genes are shown in Figure 1.3.

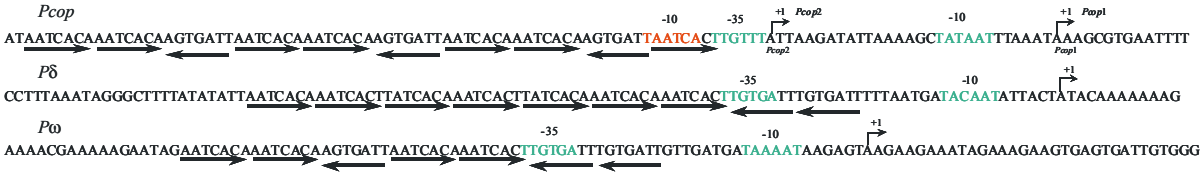

Figure 1.3: ω_2 targets in pSM19035. The conserved -35 and -10 regions of *P*cop *P*δ and *P*ω promoters are indicated in green and red. Bent arrows +1 denote known transcription start sites. The ω_2 binding sites and their relative orientations are indicated by arrows below the nucleotide sequence.

The binding of ω_2 to various arrays of heptads was examined by using site-directed mutagenesis, electrophoretic mobility shift assay (EMSA), DNase I and hydroxyl radical footprinting, surface plasmon resonance (SPR), and spectroscopic techniques (*35-36*). The apparent dissociation binding constants, $K_{d,app}$, determined by EMSA, DNase I footprinting and SPR range from ~4 to ~120 nM depending on the number of heptad and their orientation in the tested sequence. These experiments have shown that the affinity increases with the number of heptads and for tetraheptads is already comparable to the affinities of the whole ω_2 binding site composed of seven to ten heptads. The DNA sequence containing at least two heptads is required for high affinity ω_2 binding. Interestingly, EMSA indicated large differences between dissociation constants of all diheptads in the complex with ω_2. The determined $K_{d,app}$ are as follows: ⟶$_2$–ω_2 ~90 nM, ⟶⟵–ω_2 ~20 nM and ⟵⟶–ω_2 ~120 nM. The affinity of ω_2 to DNA containing only one heptad is very low with $K_{d,app}$ > 1 μM. In contrast, the Arc dimers bind to each subsite with nanomolar affinities (*37*).

The hydroxyl radical footprinting of ω_2 bound to DNA with four heptads indicated that ω_2 interacts mainly with the central 5'-TCA-3' stretch (*35*). An identical result was

obtained before by Raman spectroscopy (*36*). Genetic experiments, in which complete substitution of all the single particular base pairs was carried out in the first heptad of $\longrightarrow_2$, identified the central pentameric stretch 5'-ATCAC-3' or the complementary 5'-GTGAT-3' as essential for ω_2 binding (*35*).

In addition, CD titration experiments have shown conformational changes in DNA upon ω_2 binding (*35*), and Raman spectra indicate an induced fit of both, ω_2 and DNA, as shown by changes in vibrational modes of deoxyribose moieties and protein-induced DNA bending (*36*). This is a common feature of the operator DNA in the complex with other proteins of the MetJ/Arc family.

1.3. Arc Repressor

Arc repressor (Arc-wt) of *Salmonella typhimurium* bacteriophage P22 is involved in the switch between the lysogenic and lytic pathways of P22 by regulating the transcription of the *ant* gene during lytic growth (*38*). Phage P22 has two regulatory regions, *imm*C and *imm*I, involved in the establishment and maintenance of lysogeny. The *imm*I region encodes three proteins, antirepressor, Arc, and Mnt. The function of antirepressor is to inactivate the P22 *c*2 repressor encoded by the *imm*C region. Arc and Mnt are repressors that regulate gene expression in *imm*I region. During lytic growth of P22, Arc repressor binds to its operator site and blocks transcription of *arc* and *ant* (encoding antirepressor) genes. During lysogeny Mnt represses synthesis of both Arc and antirepressor.

Arc-wt is a small homodimeric protein and belongs to the MetJ/Arc structural superfamily. Proteins of this family form a ribbon-helix-helix DNA binding motif (*32*). In the cocrystal structure, two Arc dimers bind to adjacent subsites of a 21-bp operator DNA (Figure 1.4). An antiparallel β-sheet of Arc recognizes bases in the major groove, and N-terminal arms contribute to the interaction by direct contacts with DNA subsites (*29*). Arc-wt and Arc-F10V complexes with operator DNA form very similar structures (Figure 1.4) (*24, 29*). An extraordinary feature is observed in the crystal structure of the Arc-wt complex, namely, the protrusion of the aromatic rings of Phe10 and Phe10' and their packing against phosphate oxygens and ribose atoms of the DNA backbone (Figure 1.4) (*29*). Previous studies have shown that mutations at Phe10 reduce operator binding and protein stability (*39-40*). Furthermore, three Phe10 mutants that produced stably folded proteins with full repressor activity were probed. These studies demonstrated that Phe10 is essential for high affinity binding of Arc to its operator DNA and contributes to the

discrimination between operator and non-operator DNA (*24*). In free form, Arc-wt is the most stable Arc variant followed by Arc-F10V, Arc-F10Y, and Arc-F10H. Interestingly, Arc-F10H is the least stable free protein but it forms the most stable complex with operator DNA out of all tested Arc Phe10 mutants (*24*).

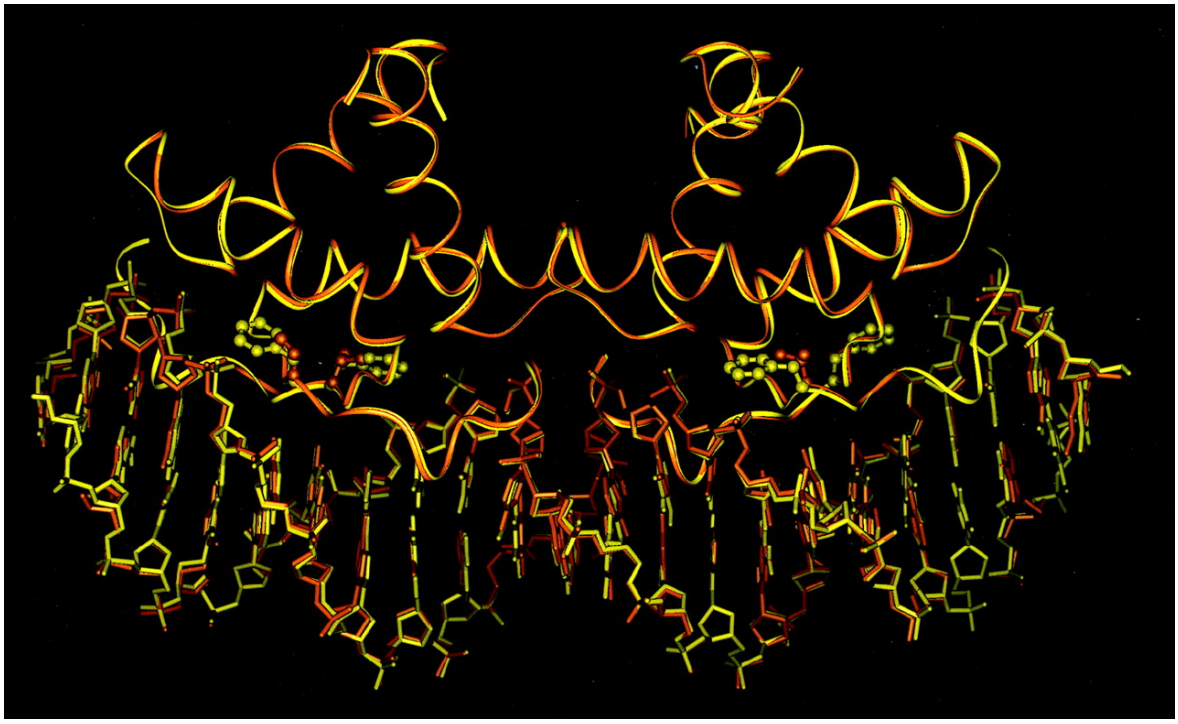

Figure 1.4: Superposition of two protein–DNA crystal structures of wild type Arc (yellow) and Arc-F10V (orange). The Val10 of Arc-F10V and Phe10 of Arc-wt are shown in ball stick representation. (Adopted from ref 24)

Solution structures of Arc-wt and Arc-MYL (with amino acid exchanges R31M/E36Y/R40L) and the dynamics of Phe10 in these proteins were studied by NMR spectroscopy (*26, 33, 41*). In the free Arc repressor dimer, at and below room temperature, Phe 10 exists predominantly in the "core" conformation where it is buried in a hydrophobic core (12 % solvent accessibility). A second, less stable "bound" conformation exists in Arc-wt and in the Arc-MYL variant with protruded Phe residues (similar to the conformation found in the Arc repressor–operator DNA complex). In the crystal structure of Arc-P8L (substitution of Pro8 by Leu) Phe10 was found only in the "core" conformation but not in the "bound" conformation (*25*).

1.4. Regulatory Protein KorB

Plasmid RP4 is a member of *Escherichia coli* incompatibility group P (IncP-1α) (*42-43*). It is a self-transmissible, broad host range, resistance plasmid of about 60 kb. IncP-1α plasmids are capable of conjugative cell-cell transfer and maintain themselves in a wide variety of Gram-negative bacteria. Due to this promiscuity they are of particular interest. Regulatory proteins KorA, KorB, KorC, IncC, and TrbA are major factors in control and coordination of replication, transfer, and partitioning functions (*44-48*) and contribute to a large extent to the survival properties of IncP-1α plasmids.

IncC belongs to the group of ParA proteins and KorB is a member of the ParB family of proteins that are encoded on plasmids and bacterial chromosomes and are involved in genome partitioning (*49-52*). Thus, KorB plays a direct role in the partitioning of plasmid RP4 and moreover, functions as a transcriptional repressor of RP4 genes.

KorB is composed of 358 amino acids (39,011 Da) and exists in purified form as a dimer in solution (*53*). O_B, the operator sequence 5'-TTTAGC($^{G}/_{C}$)GCTAAA-3' of KorB, occurs 12 times on the RP4 genome and 11 times on the related IncPβ plasmid R751. The 12 O_B sites were classified according to their positions relative to RP4 promoters in three classes. Class I sites are located 39/40 bp upstream of a transcription start site, class II sites map further upstream or downstream of promoters within 80-190 bp of a transcription start site. KorB represses promoters carrying these O_B sites (*45-46, 54-56*). Class III O_B sites are more than 1 kb away from any known promoter, and whether KorB has effects on these sites has not been elucidated. KorB acts cooperatively with KorA in transcriptional repression of the *kilA, trfA* and *korAB* operons. KorB is also involved in the negative control of *kilB* operons. According to their affinity to KorB, the 12 O_B sites fall in three groups. Protein IncC1 enhanced binding of KorB to 11 of the 12 O_B sites except O_B3. It was suggested that IncC1 influences the multimeric state of KorB and thus its binding to O_B DNA and flanking sequences.

The KorB monomer is composed of two domains, DNA binding domain and KorB-C. Clones were constructed expressing KorB-O (contains the DNA binding domain) and KorB-C separately, and the 3D structure of KorB-C, consisting of 62 amino acids (residues K297-G358), has been elucidated (*43*). KorB-C is mainly responsible for the dimerization of the protein to facilitate operator binding, whereas KorB-O represents its DNA-recognition and binding domain. The structure of KorB-O–operator DNA complex has been recently solved (Figure 1.5) (*57*).

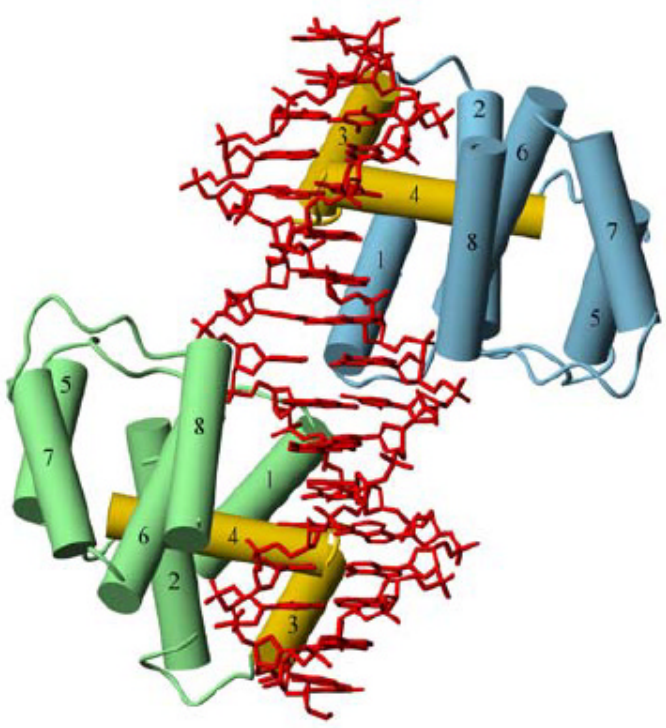

Figure 1.5: Molecular structure of KorB-O–operator complex. Two KorB-O monomers (blue and green) are occupying the half sites of O_B. The α-helices are illustrated as cylinders. Helices 3 and 4, drawn in yellow, form the HTH binding motif. (Adopted from ref 57)

KorB-O consists of 177 amino acids (residues R117-K294 of the full KorB sequence) and is composed of 8 a-helices. KorB-O forms helix-turn-helix (HTH) binding motif with helices α3 and α4 directly involved in contacts with DNA major groove as observed in other bacterial repressor proteins. Additional amino acids involved in operator binding are harboured in helices α6 and α8. Three hydrogen bonds are formed between operator DNA and KorB-O (*57*). Thr211 of helix6 makes a hydrogen bond with N4 of Cytosine11; and Arg240 of helix8 forms two hydrogen bonds with N7 and O6 of Guanine10. The experiments with site-specific mutations of these two residues have shown that the residues are essential for specific O_B binding (*57*). Interestingly, KorB HTH motif is not involved in specific DNA binding and makes only several non-specific contacts with DNA backbone.

1.5. DNA–Binding Protein Sac7d

Complex formation with DNA-binding proteins drives DNA superhelicity, which in turn regulates the double helix stability. DNA in eubacteria and eukarya is overall negatively supercoiled due to gyrase action and wrapped around the histone core of the

nucleosome. In contrast, episomal DNAs in thermophilic archaea are relaxed to positively supercoiled forms (*58-59*). In archaea the mechanism for organizing DNA into a compact form is not well understood. Members of the Sso7d/Sac7d protein family and other related proteins are believed to play an important role in DNA packaging and maintenance. Sso7d/Sac7d are small, abundant, basic and non-specific DNA-binding proteins of the hyperthermophilic archeon *Sulfolobus* (*60-61*).

The Archeon *Sulfolobus acidocaldarius* grows in acidic hot springs with a growth temperature between 60 and 85 °C (*62*). The 7-kDa chromatin proteins of *S. acidocaldarius* consist of five species, designated Sac7a, b, c, d, e in order of increasing basicity (*63-65*). Sac7d (7477 kDa) and Sac7e (7338 kDa) differ by six amino acid residues and are encoded by distinct genes. Sac7a and Sac7b are C-terminal truncated forms of Sac7d. The similar Archeon *S. solfataricus* grows optimally at temperatures 75-85 °C. Only one chromatin protein of this species has been characterized (*66*). It is referred to as Sso7d (7019 kDa) based on homology with Sac7d.

These chromatin proteins unfold reversibly over a wide range of temperature, pH, and salt concentration. The native Sac7 proteins are remarkably thermostable with an unfolding midpoint at pH 7 near 100 °C (*61*). The binding of Sso7d/Sac7d to DNA is non-cooperative with micromolar affinity, without strong sequence preference. The T_m values of Sac7–DNA complexes are by ~40 °C higher than those of free DNA (*67*). The structures of several Sac7d/Sso7d–DNA complexes have been studied by X-ray crystallography (*68-70*), NMR (*71*) and low-angle X-ray scattering (*72*). Figure 1 shows a model structure of the Sac7d–DNA complex with an octamer $d(GCGATCGC)_2$ (*68*).

In the protein–DNA complexes, the protein structures are similar to those of the free proteins (*60, 73*), consisting of an incomplete β-barrel made of a triple-stranded β-sheet orthogonal to a β-hairpin. The small β-barrel is capped by an amphiphilic C-terminal α-helix. The triple-stranded β-sheet is placed across the DNA minor groove. The side chains of Val26 and Met29 intercalate into DNA base-pairs and cause a sharp single step kink (~60 °) in the DNA duplex (*68-70*). The DNA distortion is associated with helix unwinding (~ –12 °), making a total bend of 72°. The double stranded DNA is adapted to the irregular surface of Sac7d by numerous changes in its conformation. Many nucleotides surrounding the wedge site adopt the less common C3´-endo (N type) sugar puckers.

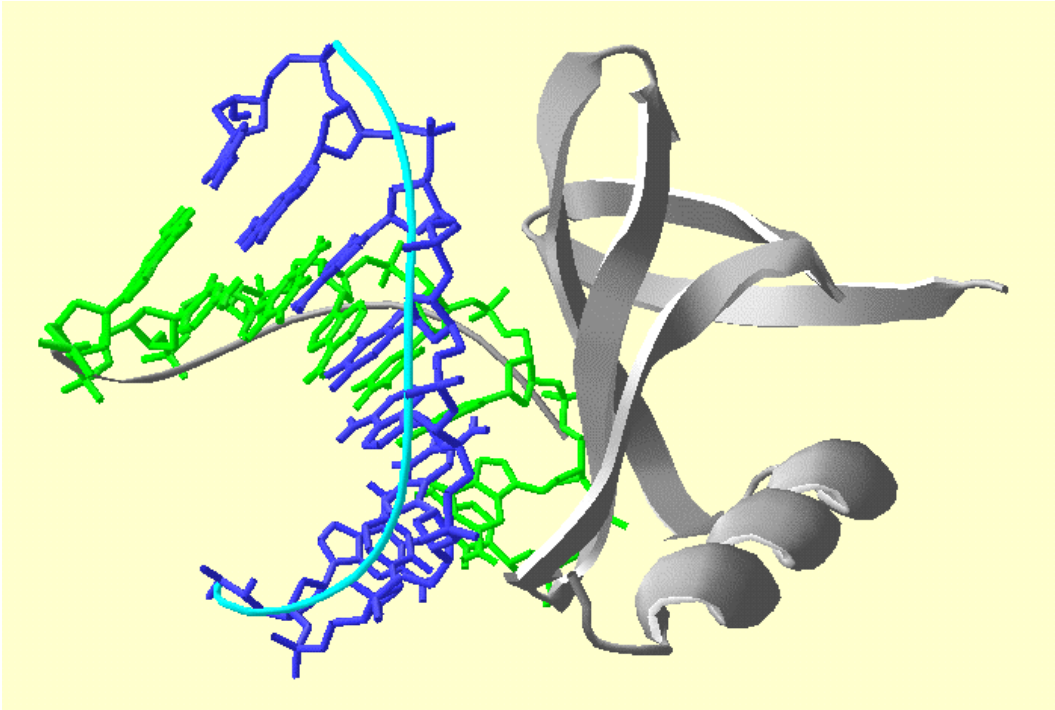

Figure 1.6: Model of the Sac7d–d(GCGATCGC)$_2$ complex. The figure was prepared using SPDBV (DeepView/Swiss-PdbViewer from http://www.expasy.org/spdb/) and the atomic coordinates from pdb file 1AZP.

2. Objectives

In the presented work will be shown possibilities and limitations of the Raman spectroscopy on studies of protein–DNA interactions. Three transcription factors (Omega protein, Arc repressor, and KorB) will be investigated of which two (Omega and KorB) act as global gene regulators and one (Sac7d) is participating in the organization of DNA in Archea.

Bacterial, plasmid-encoded protein Omega regulates the transcription of genes required for the control of plasmid copy number and their stable maintenance. Omega protein binds specifically to DNA containing at least two consecutive copies of heptads with the consensus sequence 5'- $^{A}/_{T}$ATCAC$^{A}/_{T}$ -3'. The heptads can be arranged in the operator regions in direct or inverted orientation. The solution properties of Omega protein, several Omega variants, operator DNA models containing one, two, and four heptad units in the sequence, and Omega-DNA complexes will be characterized.

KorB is encoded by the *korB* gene of plasmid RP4 and regulates the transcription of RP4 genes which are involved in replication, transfer, and stable inheritance of RP4 by binding to a 13-base pairs long palindromic sequence O_B (5' TTTAGC($^{G}/_{C}$)GCTAAA 3'). KorB has a molecular mass of 39 kDa and is composed from two domains of approximately equal size. The C-terminal domain participates in the dimerization of KorB monomers, and the N-terminal domain includes the DNA binding site. Solution properties of wild type KorB, separated N- and C-domains and complexes of wild type KorB and recombinant KorB-O with a 17 bp operator DNA model oligonucleotide will be investigated. KorB-O is composed of a part of the N-terminal domain of KorB with conserved ability to bind DNA.

A similar approach will be applied in studies on transcriptional repressor Arc. Arc repressor of *Salmonella* bacteriophage P22 is involved in the switch between the lysogenic and lytic pathways of P22 by regulating the transcription of the *ant* gene during lytic growth. Arc is a small homodimeric protein and belongs to the MetJ/Arc structural superfamily. Proteins of this family form a ribbon-helix-helix DNA binding motif. Solution properties of wild type Arc repressor and mutant Arc-F10H, as well as their complexes with operator DNA will be analyzed. Phe10 directly participates in the interaction with DNA, therefore from comparative studies important aspects of the interactions may be elucidated.

The Sac7d protein of the thermophilic Archeon *Sulfolobus acidocaldarius* is believed to play an important role in DNA packaging and maintenance. Sac7d is a small, abundant, basic and non-specific DNA-binding protein. When bound to DNA it increases the melting temperature T_m of DNA by ~40 °C. Solution properties of Sac7d and its complex with a GC rich DNA oligonucleotide [sequence $d(GAGGCGCCTC)_2$] will be characterized.

Three dimensional structures of KorB–, Sac7d–, and Arc repressor–DNA complexes are known. The aim of the Raman measurements is the following: First, the Raman spectra of complexes of known structures will be compared with existing protein–DNA models to evaluate the protein–DNA contacts, the DNA backbone conformations, the surroundings of amino acid side chains, and the secondary structure of the proteins. Second, the known crystal and NMR models will be used to correlate the structural information and Raman difference bands, thus improving the spectroscopic knowledge. This information will be than applied in the characterization of the Omega–DNA interaction where no complex structure is available so far.

From studies on these systems we expect contributions to a deeper understanding into mechanisms of protein–DNA recognition, and in addition the results will provide an extension of the methodological possibilities of Raman spectroscopy in this field.