Structural and functional studies of the global transcription regulator KorA of plasmid RP4

Dissertation zur Erlangung des akademischen Grades des

Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie

der Freien Universität Berlin

vorgelegt von

Dipl.-Chem. Bettina König

aus Berlin

Mai, 2009

1. Gutachter: Prof. Dr. Udo Heinemann

2. Gutachter: Prof. Dr. Markus Wahl

Mündliche Prüfung am: 07. Oktober 2009

Selbstständigkeitserklärung

Diese Arbeit wurde im Zeitraum von Oktober 2004 bis Mai 2009 in der Forschungsgruppe Makromolekulare Strukturen und Interaktionen am Max-Delbrück-Centrum für Molekulare Medizin in Berlin-Buch unter der Leitung von Prof. Dr. Udo Heinemann angefertigt.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und auf Grundlage der angegebenen Hilfsmittel angefertigt habe.

Teile dieser Arbeit sind Bestandteil folgender Publikation:

König, B., Müller, J.J., Lanka, E., Heinemann, U. (2009) Crystal structure of KorA bound to operator DNA: insight into repressor cooperation in RP4 gene regulation. *Nucleic Acids Res*, **37**, 1915-1924; doi: 10.1093/nar/gkp044.

Table of contents

1 Introduction 1
1.1 Transcription regulation in bacteria 1
1.1.1 Transcription initiation 1
1.1.2 Transcription factors 2
1.1.3 DNA binding of transcription factors 5
1.1.4 Cooperativity between transcription factors 6
1.2 Plasmids 7
1.2.1 Host range 7
1.2.2 Classification 7
1.2.3 Survival strategies 8
1.2.4 The RP4 plasmid 9
1.2.5 The *korAB* operon 11
1.2.6 KorA 12
1.2.7 Cooperativity with KorA 13
1.2.8 The aim of this study 13
2 Materials 15
2.1 Instruments 15
2.2 Chemicals 17
2.3 Strains and plasmids 18
2.4 Media and buffers 19
2.5 Synthetic oligonucleotides 25
2.5.1 Primers 25
2.5.2 Oligonucleotides for QuikChange mutagenesis 26
2.5.3 Oligonucleotides for ITC and crystallization 26
3 Methods 27
3.1 Molecular biology methods 27
3.1.1 Polymerase chain reaction (PCR) 27
3.1.2 Gel electrophoresis 28
3.1.3 Annealing of DNA oligonucleotides 28
3.1.4 Determination of DNA concentration 29
3.1.5 QuikChange mutagenesis (QCM) 29
3.1.6 Plasmid preparations 31
3.1.7 Transformation of chemically competent *E. coli* cells 31

3.2 Protein purification and characterization methods 32
3.2.1 Expression 32
3.2.2 Cell lysis 32
3.2.3 Precipitation with polyethyleneimine 32
3.2.4 Ammonium sulfate precipitation (ASP) 32
3.2.5 Dialysis 33
3.2.6 Affinity chromatography 33
3.2.7 Cation-exchange chromatography 34
3.2.8 Gel-filtration chromatography 34
3.2.9 Determination of protein concentration 35
3.2.10 SDS and native PAGE 35
3.2.11 Electrophoretic mobility shift assay (EMSA) 36
3.2.12 Gel staining procedures 37
3.2.13 Isothermal titration calorimetry (ITC) 37
3.2.14 Analytical ultracentrifugation 40
3.3 Structural methods 42
3.3.1 Crystallization 42
3.3.2 Diffraction data collection 44
3.3.3 Data processing 46
3.3.4 The electron density 47
3.3.5 Obtaining phases 48
3.3.6 Model building and refinement 52
4 Results 55
4.1 Structure determination 55
4.1.1 Recombinant expression of *korA* 55
4.1.2 Purification of KorA 55
4.1.3 Formation of the KorA-O_A* complex 58
4.1.4 Crystallization of the KorA-O_A* complex 60
4.1.5 Data collection 60
4.1.6 Structure analysis and refinement 62
4.2 The structure of the KorA-O_A* complex 65
4.2.1 The KorA monomer 65
4.2.2 The KorA dimer 66
4.2.3 The DNA-binding domain of KorA 68
4.2.4 Orientation of the operator site 69
4.2.5 The operator binding site of KorA 70
4.2.9 The conformation of the operator site 71

5 Discussion 72
5.1 Success and failure of experimental MAD phasing....72
5.2 Specific DNA-binding of KorA 74
5.3 Structural homologs of KorA 77
5.4 KorA homologs of other IncP-1 plasmids 78
5.5 The dimerization module....79
5.6 Cooperation mechanism in RP4....82
5.7 Conclusion and closing remarks....84
6 Summary 86
7 Zusammenfassung 88
Appendix A: References 90
Appendix B: Directories 102
Appendix C: Abbreviations....104
Curriculum vitae....107
Danksagung 108

1 Introduction

1.1 Transcription regulation in bacteria

Bacteria have evolved very sophisticated responses to adapt to dramatically and instantly changing environmental conditions. Rapid adaptation to a given stress or change in nutrient source is achieved by switching on and off the expression of specific genes. Bacteria are very efficient at increasing the production of certain proteins when they are needed, or to shut it off if metabolically and energetically wasteful. Hence, bacteria have developed a wide range of mechanisms to regulate gene expression. Every step from transcription initiation to protein turnover can be tightly controlled; however, regulation at the stage of transcription initiation is the dominant form of regulation (Rojo 1999).

1.1.1 Transcription initiation

The central element in transcriptional regulation in bacteria is the DNA-dependent RNA polymerase (RNAP), which is composed of several subunits, shown in Figure 1.1 (Ebright 2000; Murakami *et al.* 2002; Browning and Busby 2004). The core enzyme ($\alpha_2\beta\beta'$) is responsible for elongation of the transcript, whereas the holoenzyme ($\alpha_2\beta\beta'\sigma$) contains an additional sigma (σ) factor (Burgess *et al.* 1969; Helmann and Chamberlin 1988), which ensures the recognition of specific promoter sequences. The σ subunit positions the RNA polymerase holoenzyme at a target promoter and facilitates unwinding of the DNA duplex near the transcript start site (Wösten 1998). Bacteria contain several σ factors, each one directing RNAP to a specific set of promoters (Helmann and Chamberlin 1988). Domains 2 and 4 of the RNA polymerase σ subunit recognize the promoter -10 and -35 elements, respectively. Domain 3, on the other hand, specifically binds to the extended -10 element, a 3-4 bp motif located immediately upstream of the -10 hexamer (Murakami *et al.* 2002; Sanderson *et al.* 2003). Another important promoter element, the UP element, is recognized by the C-terminal domains of the RNAP α subunits (Ross *et al.* 2001). The specific binding of RNAP to the promoter is therefore exquisitely adapted to the promoter characteristics, resulting in unequal distribution of RNAP between different transcription units. This helps bacteria to carefully regulate the binding of RNAP across their chromosome, and, because RNAP is in short supply, the competition between promoter sites is enormous.

Transcription starts with the initial binding of RNAP to the promoter resulting in a closed complex. Subsequently, local melting of DNA strands at the transcription start point leads to the formation of an open complex (deHaseth *et al.* 1998). The synthesis of the DNA template-directed RNA chain begins. The initiation phase is finished, when RNAP breaks contacts with the promoter, releasing the sigma factor. RNAP is moved into a productive elongation complex, which is responsible for RNA chain extension (Spassky *et al.* 1985; Lloyd *et al.* 2001).

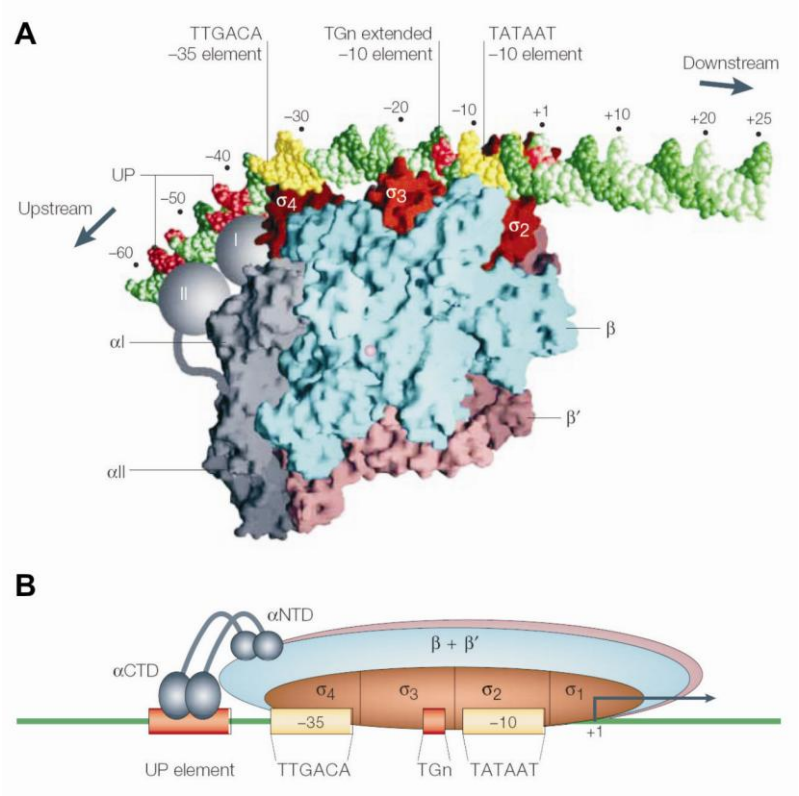


Figure 1.1: RNA polymerase bound to a promoter site. (**A**) Model of the initial binding of RNAP to a promoter (closed complex), based on crystallographic studies by Murakami *et al.* (2002). DNA, shown in green, with the -10 and -35 elements and the extended -10 and the UP elements shown in yellow and red, respectively. The RNAP is shown with its β and β' subunits (light blue and pink), the αNTDs (gray), and the different sigma domains (red). The αCTDs (gray spheres) bind to the UP element. A Mg^{2+} ion bound to the active site is shown in magenta. (**B**) A model of RNAP bound to a promoter by Browning and Busby (2004).

1.1.2 Transcription factors

Transcription factors coordinate the expression of genes in response to environmental signals. Some are global regulators controlling operons that belong to different metabolic pathways, whereas local regulators manage just one or two genes (Gottesman 1984). To control their actions, the cell can employ different methods. The DNA-binding affinity of transcription factors can be altered by small ligands, and their activity can be adjusted by covalent modifications or by concentration dependencies. Also, other regulatory proteins to which they bind cooperatively can be kept occupied.

Each step in transcription initiation can, in theory, be regulated by transcription factors interacting at or near the corresponding promoter. Most prokaryotic transcription factors are DNA-binding proteins, and their promoter specificity depends on preferential binding

to a specific DNA sequence, the operator binding site. Depending on the nature of the regulatory protein and the relative position of its binding site in the upstream DNA regions of transcription regulation, it either activates or represses the transcription of the gene (Collado-Vides *et al.* 1991; Gralla 1996). Several transcription factors can also function as either repressor or activator according to the target promoter (Pérez-Rueda and Collado-Vides 2000).

Many promoters lack a good match to the consensus elements for RNA polymerase binding and require an additional protein, known as transcription activator, to function. There are three common mechanisms of how activation is achieved, as shown in Figure 1.2. First, the activator protein binds upstream of the -35 element and interacts with RNAP *via* its αCTD. This cooperation guides RNAP to the exact position where it can bind to the promoter site. Second, the activator protein binds near the -35 element and then interacts with the sigma 4 domain of RNAP. Again, this cooperation between regulator and RNAP assists in binding to the promoter site easier. Modification of the DNA conformation at the promoter site by the bound activator is the third mechanism. Deformation of DNA can influence the interaction of RNAP and promoter site.

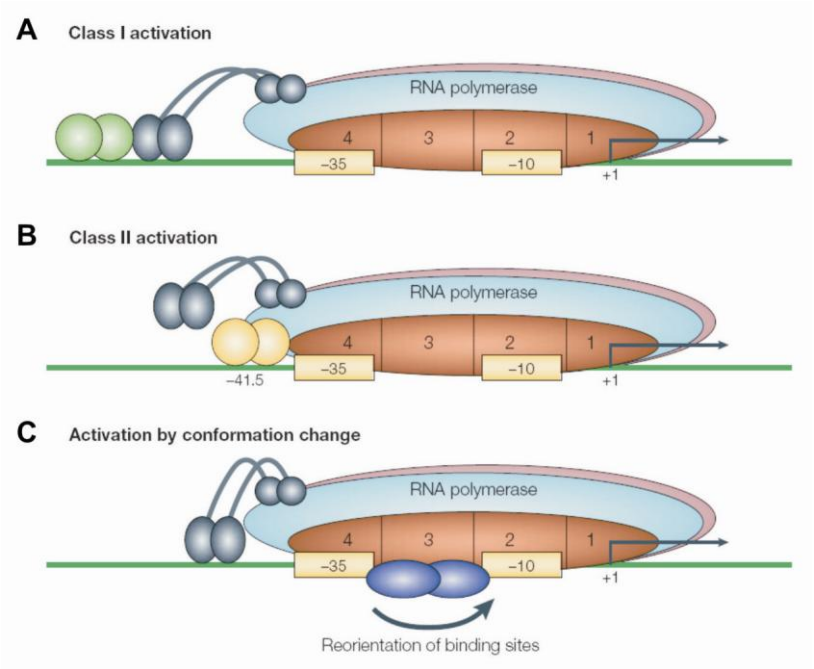


Figure 1.2: Mechanisms for transcription activation. (**A**) The dimeric activator (green) is bound further upstream of the promoter site. By interacting with the αCTD of RNAP (gray) the sigma domains of RNAP (brown) can bind more easily to the promoter. (**B**) The activator (yellow) binding site is situated close to the -35 element. The activator can directly bind to the sigma 4 domain directing the RNAP to the promoter. (**C**) Conformational changes of the promoter site can enhance RNAP binding to the promoter. The activator (blue) binds at or near the promoter and reorients the -10 and -35 element (Browning and Busby 2004).

Some promoters are active in the absence of additional factors and, when the genes under their control are not required, they are silenced by transcription repressors. Repression efficiency depends to a large extent on the competition of RNAP with the repressor for their overlapping binding sites and on the rate of promoter clearance (Lanzer and Bujard 1988). Three basic mechanisms are used for repression, see Figure 1.3. The first is interference with RNAP binding to the promoter, where the repressor binding site is placed within or close to the promoter elements. The second mechanism involves DNA-looping, where at least two repressor proteins bind to promoter-distal sites and completely block the promoter site for RNAP binding. The third mechanism describes the repression of an activator protein by direct interaction.

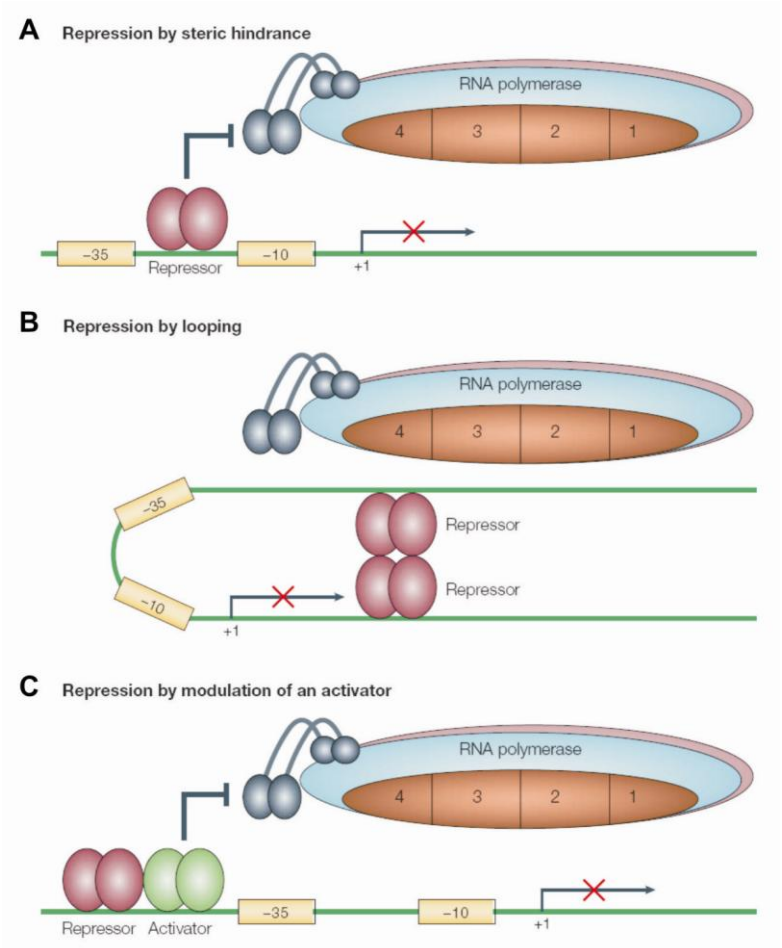


Figure 1.3: Mechanisms for transcription repression. (**A**) The repressor interferes with RNAP recognition of the promoter site by blocking it. (**B**) Repressors bind to distal sites, and interaction results in DNA looping, which blocks the promoter site. (**C**) The repressor binds to an activator (green) and disturbs its function (Browning and Busby 2004).

1.1.3 DNA binding of transcription factors

Three frequently used DNA-binding motifs of prokaryotic regulatory proteins have been identified: the helix-turn-helix (HTH) motif, the winged-helix motif and the β-ribbon motif (Pabo and Sauer 1992; Huffman and Brennan 2002). Generally, these motifs use either α helices or β sheets to dock into the major groove of B-DNA. This groove allows to create twice as many specific contacts with its bases as the minor groove does (Seeman *et al.* 1976). Its three-dimensional appearance and electrostatic environment define the shape of the protein's secondary structure elements that fit into it. The α helix is most frequently used, because it is shape-complementary to the major groove, providing the highest density of possible contacts with the bases and the backbone of the DNA.

In prokaryotes, the HTH motif is the most widely used DNA-binding motif. It consists of approximately 20 amino acids forming two α helices connected by a short turn. The two α helices, the C-terminal recognition helix and the N-terminal scaffold helix, are held at a fixed angle. The relative orientation between the helices is well conserved in all DNA-binding domains (Suzuki and Brenner 1995). The HTH motif is no separate stable unit and cannot fold independently. Therefore, a generally preceding third α helix stabilizes the motif to function as a compact, globular domain. The HTH motif is usually applied more than once to recognize symmetric operator sequences. Hence, proteins adopt the matching canonical fold by forming homodimers, heterodimers, or even tandem recognition motifs within a single polypeptide (Brennan and Matthews 1989; Harrison 1991; Wintjens and Rooman 1996; Gallegos *et al.* 1997).

Transcription regulators are complementary to the surface features of their specific DNA binding site. The recognition helix easily penetrates the major groove, where protein side chains anchored in the α helix form contacts with the bases and with the DNA backbone (Pabo and Sauer 1992). A large number of contacts are formed with the DNA, involving hydrogen bonds, ionic and hydrophobic interactions. The many contacts formed at the protein-DNA interface ensure strong and highly specific interactions. Crucial contacts are usually made by the protein side chains of the HTH motif by hydrogen bonds, but also outside of HTH motif important contacts with the DNA are made helping to fine-tune the interactions. Another feature recognized by regulatory proteins is the overall geometry or deformability of the double helix that depends entirely on its nucleotide composition. The fit between protein and DNA must be tight to achieve strong binding, and often distortions in DNA conformation increase the complementarity. Besides the relative position of the operator binding site to the promoter (Collado-Vides *et al.* 1991; Gralla 1996), the relative position of the HTH motif within the protein seems to be important as well. Repressors

have the HTH motif in general at the N terminus, whereas activators carry it close to the C-terminus (Pérez-Rueda *et al.* 1998).

1.1.4 Cooperativity between transcription factors

The cooperative assembly of protein-DNA complexes is widespread in eukaryotic transcription regulation (Valentine *et al.* 1998; Merika and Thanos 2001), while in prokaryotes only a few examples are found (Barnard *et al.* 2004). Cooperativity occurs, when many factors are connected in response to one another to increase the speed, the affinity and the specificity of protein-DNA binding, and thereby affect the activity of RNAP. This leads to increased activation or repression of a promoter by direct protein-protein interaction or modulation of local DNA conformation.

In prokaryotes, there are two different approaches of cooperativity between two dimeric regulatory proteins and one DNA molecule (Figure 1.4). First, the transcription regulators bind separately to their adjacent operator sites. This generates a local concentration of both proteins that is high enough to support stable protein-protein interactions. The second and more common approach is that the regulators form tetramers before binding to the DNA simultaneously. Here, one dimer anchors the protein complex to its operator site near the second binding site. Thus, the second regulator is being recruited to its own binding site (Gralla 1989; Senear *et al.* 1998). The alignment and distance of the two operator sites as well as the shape and deformability of the DNA binding site can affect cooperativity to a great extent. Very distant operator sites can only come into close contact, when the DNA adopts the optimal conformation (DNA looping), and when the protein-protein interactions are very strong (Brenowitz *et al.* 1991).

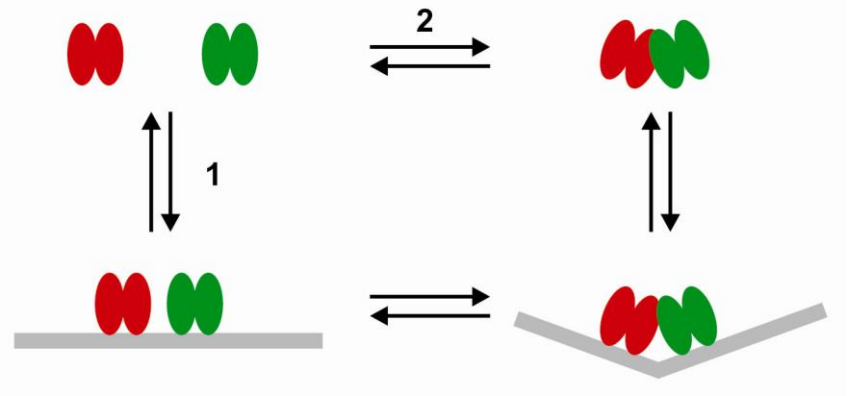


Figure 1.4: Cooperativity between transcription regulators. The formation of a protein-DNA complex between two dimeric regulatory proteins shown in red and green, respectively, binding cooperatively to adjacent DNA sites colored in gray, can occur in two different ways. (**1**) The dimers may bind DNA separately followed by protein-protein interactions, and often conformational changes in DNA as well. (**2**) The dimers oligomerize before binding simultaneously to DNA. Adapted from Senear *et al.* (1998).

1.2 Plasmids

Bacteria contain small extrachromosomal DNA molecules in addition to a large and complex chromosome. These molecules are called extrachromosomal replicons or plasmids (Lederberg 1952; 1998). Plasmids exist in both circular and linear forms. Linear mitochondrial plasmids are present in many fungi and in some plants, but they seem to be absent from most animal cells. Plasmids display an amazing diversity of characteristics, such as size, modes of replication and transmission, host ranges, and the repertoire of genes that they carry. They have adopted a variety of strategies to ensure their own faithful replication, maintenance, and transfer. They also impart a wide assortment of phenotypes to the cells that harbor them. Despite their small size, plasmid replication and gene expression impose a metabolic burden on the host cell. Therefore, in contrast to the bacterial chromosome, plasmids have evolved systems to guard DNA stability and integrity, rather than specializing in cell survival under nutritional starvation.

1.2.1 Host range

Plasmids rely on self-encoded as well as host-encoded factors for duplication of their genetic material. Therefore, the host range they can stably exist in depends on the ability of the plasmid replication apparatus to interact with the enzymatic machinery of its host (del Solar *et al.* 1996). Some naturally occurring plasmids can only be stably maintained in a single or closely related bacterial host (narrow host range) while others have the ability to replicate in unrelated bacterial species (broad host range or promiscuous). Broad-host-range plasmids have been investigated over the years for many reasons, including their use as cloning vectors for biotechnological purposes, their role as agents of horizontal gene transfer in microbial communities, and the insight they provide into the mechanisms of initiation of DNA replication.

1.2.2 Classification

A plasmid classification system based on incompatibility was developed in the early 1970s and is now commonly used (Datta and Hedges 1971). Incompatibility is a manifestation of relatedness in the sharing of common elements which are involved in plasmid replication and partitioning (Novick 1987; Austin and Nordström 1990). Plasmids with closely related replication control systems are mostly unable to be maintained stably in the same host cell. These plasmids are incompatible, and they are assigned to the same incompatibility (Inc) group.

For *Pseudomonas aeruginosa*, for example, 14 incompatibility groups are currently recognized (Table 1.1). The broad host range of some *Pseudomonas* plasmids (Olsen and Shipley 1973) has resulted in the classification of plasmids such as RP4 into the P-1 group in *Pseudomonas* and into the P group in *Enterobacteriaceae*.

Inc group[a]	Plasmid[b]	Phenotype[c]	Size (kb)
P-1	RP4	Cb Km Tc Te	60
P-2	pMG1	Gm Sm Su Bor Hg Te Phi	481
P-3	RIP64	Cb Cm Gm Su Tm Hg	147
P-4	R1162	Sm Su	8
P-5	Rms163	Cm Su Tc Bor	224
P-6	Rms149	Cb Gm Sm Su Phi	56
P-7	Rms148	Sm Phi	224
P-8	Fp2	Hg Pmr Cma	93
P-9	R2	Cb Km Sm Su	68
P-10	R91	Cb	54
P-11	pMG39	Cb Gm Km Sm Su Tm	93
P-12	R716	Sm Hg	170
P-13	pMG25	Cb Cm Gm Km Sm Su Tm Bor	102
P-14	pBS222	Tc	17

[a]Plasmids within the same incompatibility (Inc) group are unable to coexist in the same host cell.
[b]A single plasmid was chosen to represent each Inc group.
[c]Phenotypes are designated to plasmid-mediated resistance (see also Appendix C).

Table 1.1: Plasmid incompatibility groups in *P. aeruginosa* (Taylor and Bradley 1987).

1.2.3 Survival strategies

Plasmids do not carry genes which are essential for growth of their host under non-stressed conditions but often adaptive genes which are necessary for bacterial survival in a special environment. Therefore, plasmids are a very important addition to the bacterial genome. Many attempts to cure a strain of its plasmid have proven unsuccessful. Such high stability depends on a wide assortment of phenotypes, the ability of horizontal DNA transfer, an efficient replication system with a copy-control mechanism that ensures a reasonable even plasmid copy number from cell to cell and mechanisms that ensure efficient segregation. However, plasmid survival is not dependent on a single strategy, but rather on additive effects.

Plasmids can mediate resistance and thus confer growth advantages to their hosts in an environment where antibiotics are present. They have an enormous ability to acquire drug resistance determinants from a variety of sources and to constantly evolve new ones (Tran and Jacoby 2002). Promiscuous plasmids are carriers in the spread of antibiotic resistance to new hosts. During conjugation, plasmid DNA is directly transferred between cells that have come into contact with one another.

To prevent elimination, plasmids have to maintain their characteristic copy numbers within growing cell populations. Fluctuation of copy numbers is unavoidable and depends on the host and growth conditions. Special control systems prevent a large decrease or increase of copy numbers (del Solar and Espinosa 2000). While low copy numbers can lead to plasmid loss (Nordström *et al.* 1984; Wagner and Brantl 1998), high copy numbers may lead to runaway replication, which kills the host cell after approximately four generations (Uhlin and Nordström 1978).

High-copy-number plasmids segregate randomly, but the probability of generating a plasmid-free cell is exceedingly small (Summers and Sherratt 1984; Summers *et al.* 1993). Low-copy-number plasmids minimize the metabolic burden on the host cell. It is therefore of great benefit for the plasmid to tightly regulate the replication events. Despite its benefits for plasmid survival, a low copy number can reduce the chances of transmission to daughter cells. Therefore, the own faithful segregation during cell division needs to be ensured. To actively dictate the specific localization of plasmids inside the bacterial cell and coordinate this localization with the bacterial cell cycle, plasmids have evolved partition systems (Gerdes *et al.* 2000).

1.2.4 The RP4 plasmid

Plasmids of the incompatibility group IncP-1 (Table 1.1) are capable of transfer and stable inheritance in most Gram-negative bacteria (Adamczyk and Jagura-Burdzy 2003). Besides various bacterial hosts they can also transfer to yeast (Heinemann and Sprague 1989) and some higher eukaryotic cells (Waters 2001). IncP-1 plasmids are divided into several subgroups, α, β, γ, and δ (Dennis 2005; Bahl *et al.* 2007). The best studied family members share a common backbone and fall in the α and β subgroups (Pansegrau and Lanka 1987): RP4 (Figure 1.5; indistinguishable from RK2, R18, R68 and RP1) for IncP-1α (Pansegrau *et al.* 1994) and R751 for IncP-1β (Thorsted *et al.* 1998). Plasmid RP4 (60,099 bp with four to seven copies per chromosome) was isolated from clinical strains of antibiotic-resistant *Klebsiella aerogenes* and *Pseudomonas aeruginosa* in 1969 at the Birmingham Accident Hospital (Lowbury *et al.* 1969). The structure and replication

properties of RP4 were of particular interest soon after *in vitro* recombinant DNA techniques were introduced.

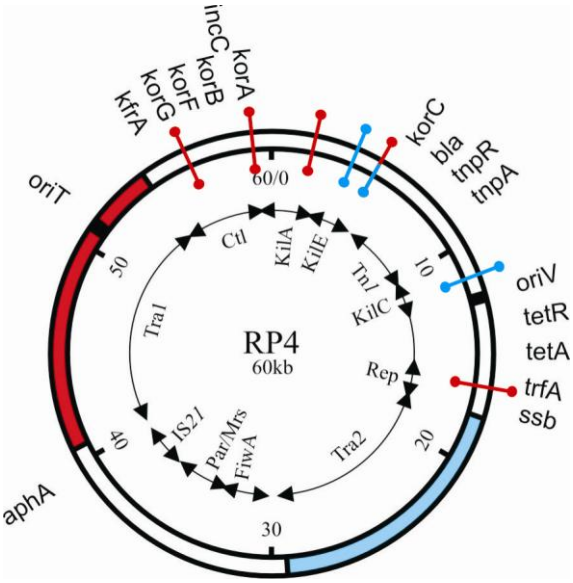


Figure 1.5: Map of plasmid RP4. The outer circle shows selected genetic loci, and the inner circle shows blocks of related genes or distinct genetic elements. The global regulatory protein KorA binds to seven operator sites, O_A1 to O_A7, on RP4. The binding sites are represented by dumbbells. Each RP4 operon is regulated by more than one repressor. The O_A site is marked in red for cooperative repression with KorB, blue for cooperative repression with KorC, and red/blue for cooperativity with both. Data were adapted from Balzer *et al.* (1992) and Pansegrau *et al.* (1994).

A fundamental regulatory strategy of RP4 is the autogenous control of expression of its survival functions. Each operon encodes at least one autorepressor, which ensures that gene expression is downregulated once a functional level of protein concentration has been achieved. In addition, the regulation of plasmid replication, transfer and stable maintenance genes is coordinated stringently by grouping them together in one operon or by arranging their respective operons close together (Bingle and Thomas 2001). Replication initiation at the vegetative replication origin (*oriV*) depends on *trfA*, and is active in a wide range of bacteria (Schmidhauser and Helinski 1985; Caspi *et al.* 2001). RP4 stabilization depends on three plasmid loci: *kilE*, *par*, and *incC/korB*. The *kilE* locus is needed for stable inheritance specifically in *P. aeruginosa* (Wilson *et al.* 1997). The *par* locus encodes a post-segregational toxicity system in the *parDE* operon (Roberts *et al.* 1994), and a multimer resolution system in the *parCBA* operon (Easter *et al.* 1998). The *incC/korB* locus codes for an active partition system (Rosche *et al.* 2000). The RP4 regulatory circuit is controlled by the global repressor proteins KorA, KorB, KorC and TrbA, each repressing several operons, while each operon is regulated by more than one repressor (Bingle and Thomas 2001).

1.2.5 The *korAB* operon

The stable maintenance of low-copy plasmids like RP4 is based on accurate DNA segregation before cell division. To ensure the equal distribution of genetic material among the daughter cells, partition (par) systems need three components: a *cis*-acting centromere-like site, and two *trans*-acting factors, a DNA-binding protein and an ATPase. To form the nucleoprotein complex that mediates genome segregation, the DNA-binding protein binds specifically to the centromere-like site and interacts with the ATPase, whose activity is essential for the movement and localization of the segrosome (Ebersbach and Gerdes 2005; Ghosh *et al.* 2006; Hayes and Barilla 2006).

The active par system of RP4 is encoded in the complex *korAB* operon (Figure 1.6). It carries the genes of its transcriptional repressors KorA and KorB (Pansegrau *et al.* 1994) as well as the gene responsible for incompatibility, *incC* (Meyer and Hinds 1982). IncC is made in two forms by two different translational starts in the same reading frame (Williams *et al.* 1998). The *incC1* (full-length gene product of 364 aa) and *incC2* (encoding an N-terminally truncated version of 259 aa) starting regions are overlapped by the start and stop codon of *korA*, respectively, in the second reading frame, as shown in Figure 1.6. This remarkable *korA*/*incC1*/*incC2* arrangement of overlapping genes, by tandem in-frame initiation regions or by using the coding capacity of a different reading frame is conserved throughout the P-plasmids (Pansegrau *et al.* 1994). While KorA contains an HTH motif, IncC1 can only bind non-specifically to DNA (Batt *et al.* 2009), but potentiates the repressor activity of KorB (Jagura-Burdzy *et al.* 1999). IncC2 and KorB are the ParA and ParB homologs, respectively, for active partition in RP4 (Williams *et al.* 1998; Siddique and Figurski 2002).

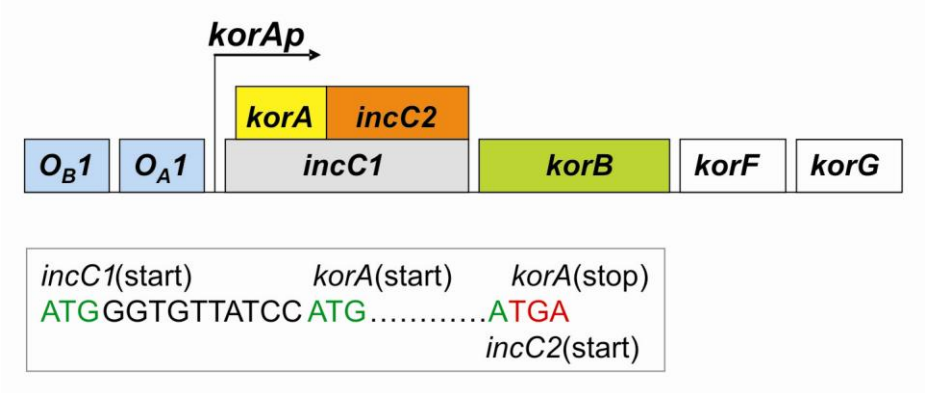


Figure 1.6: The *korAB* operon of RP4. O_A1 and O_B1 (blue) are operator binding sites for the repressor proteins KorA and KorB, respectively. The *korA*/*incC1*/*incC2* gene arrangement (yellow/gray/orange) is shown in more detail below the operon presentation. Data were adapted from Pansegrau *et al.* (1994), and Siddique and Figurski (2002).

1.2.6 KorA

KorA consists of 101 residues (11,306 Da) and exists as a dimer in solution. Each KorA monomer has two domains (Jagura-Burdzy *et al.* 1992). The N-terminal domain contains an HTH motif (Thomas and Smith 1986), and the C-terminal domain seems to have a novel dimerization motif (Bhattacharyya and Figurski 2001). KorA recognizes and binds a 12-bp symmetric operator O_A (consensus sequence 5'-GTT TAG CTA AAC-3') found at seven promoter sites on RP4 with different affinities, shown in Figure 1.5 and Table 1.2 (Jagura-Burdzy and Thomas 1995). Each site contains a 6-bp inverted repeat that is specifically bound by KorA. The flanking sequences influence the relative affinity of KorA for each site. The distance between operator and promoter site can vary. Based on this, O_A sites can be classified as defined in Table 1.2.

The operator sites fall into two classes (Table 1.2) based on the location of these sequences. Class I sites (O_A1 *korAp*, O_A3 *trfAp*, and O_A7 *klaAp*) are found 10 bp upstream and class II sites (O_A2 *kfrAp*, O_A4 *klcAp*, O_A5 *kleAp*, and O_A6 *kleCp*) are located 35 bp upstream of the transcription start point. The binding strength of KorA to operator sites within one class is very similar. The highest affinity is seen at operator site O_A1 found near the promoter of the *korAB* operon. The class II operator sites all carry one mismatch from the consensus sequence, which reduces KorA binding strength.

Operator	Promoter	Operator sequence (5' → 3')	K_{app} [nM][a]	Class[b]	ΔO_{AB}[c]	ΔO_{AC}[c]
O_A1	*korAp*[d]	AAT **GTTTAG CTAAAC** TTC	12.9	I	32 bp	-
O_A2	*kfrAp*[e]	ACT **GTTCAG CTAAAC** TCT	153.0	II	36 bp	-
O_A3	*trfAp*[f]	AAT **GTTTAG CTAAAC** TAG	20.1	I	33 bp	-
O_A4	*klcAp*[g]	ATA **GTTTAG CTAAAT** TTG	136.0	II	-	27 bp
O_A5	*kleAp*[g]	ATT **GTTTAG CTAAAT** TTC	233.0	II	92 bp	27 bp
O_A6	*kleCp*[g]	ATT **GTTTAG CTAAAA** TTG	272.0	II	-	27 bp
O_A7	*klaAp*[h]	AAA **GTTTAG CTAAAC** TTC	13.3	I	33 bp	-

[a]Apparent affinities of KorA for its operator site (Jagura-Burdzy and Thomas 1995).
[b]Class based on relative location of the O_A site with respect to promoter. Class I in the -10 region, and class II in the -35 region relative to the promoter.
[c]ΔO_{AB} and ΔO_{AC} are the center-to-center distances between the operator binding sites of KorA and KorB, as well as KorA and KorC.
Repression of KorA at the seven promoter sites was described by:
[d](Theophilus *et al.* 1985), [e](Thomas *et al.* 1990), [f](Schreiner *et al.* 1985), [g](Thomas *et al.* 1988; Larsen and Figurski 1994), and [h](Young *et al.* 1985).
The 12-bp consensus sequence is given in red and base mismatches are highlighted in gray.

Table 1.2: The seven O_A sites on RP4.

1.2.7 Cooperativity with KorA

KorA, KorB and KorC are global transcription regulators with seven, twelve and three operator sites, respectively, throughout the RP4 plasmid genome. Kor indicates the first identified function of these proteins. They are products of **k**il-**o**ver**r**ide genes, suppressing the host-lethal or -inhibitory phenotypes (KilA and KilB) (Figurski *et al.* 1985; Thomas *et al.* 1985). All KorA-regulated promoters are associated with genes for replication or stable inheritance of RP4. At five promoter regions a KorB operator site (O_B) is also present, and at three O_A sites, KorC controls repression as well, when bound to an O_C site (Figure 1.5 and Table 1.2).

KorB consists of 358 amino acids (39,011 Da) and exists as a dimer and higher multimers in solution (Balzer *et al.* 1992; Williams *et al.* 1993). The C-terminal dimerization region KorB-C with the interlocking SH3-like domains (aa 298-358) (Delbrück *et al.* 2002), and the DNA-binding domain KorB-O containing an HTH motif (aa 138-252) (Khare *et al.* 2004) are structurally known. Five O_B sites are separated from O_A center-to-center by 32 bp (*korAp*), 33 bp (*trfAp*, *klaAp*), 36 bp (*kfrAp*), or 92 bp (*kleAp*). The interaction of both proteins is important to accomplish strong repression. KorB acts very flexibly and cooperates with other repressor proteins from various distances with no particular helical position required (Bingle *et al.* 2005; Chiu *et al.* 2008). KorA interacts with KorB directly through its C-terminus (Bingle *et al.* 2008). This domain exhibits a high level of conservation between KorA and two other RP4 proteins, TrbA and KlcB (Kostelidou *et al.* 1999; Bhattacharyya and Figurski 2001).

KorC is 85 amino acids in length (9,150 Da) and consists of a putative HTH motif (Kornacki *et al.* 1993). Cooperativity between KorA and KorC takes place at three promoter regions where O_A and O_C are separated center-to-center by exactly 27 bp (*klcAp*, *kleAp*, *kleCp*) (Larsen and Figurski 1994). This conserved spacing seems to be of importance, although direct interactions between both proteins have not been reported so far.

1.2.8 The aim of this study

Regulation of gene expression plays an essential role in the maintenance of plasmids and their hosts. Over the last years many three-dimensional structures of prokaryotic transcription factors could be obtained by NMR and X-ray crystallography methods. Although the regulator proteins seem to exploit more or less recurring DNA-binding motifs, most of them have been proven to function locally at their individual operator sites.

KorA was shown to coordinate gene transcription throughout the RP4 plasmid genome. Hence, the goal of my work was to determine the high-resolution structure of KorA bound to its operator binding site and to reveal its molecular mechanisms of gene regulation. In addition, the DNA-binding properties of KorA were elucidated in isothermal titration calorimetry experiments and electrophoretic mobility shift assays. A very interesting feature in RP4 is the cooperativity between the global transcription factors, but so far very little is known about it. The unusual dimerization module of KorA obtained in the crystal structure finally provided the opportunity to propose a cooperation model between KorA and KorB.

2 Materials

2.1 Instruments

Analytical scale	MR 2002	Sartorius
Precission scale	MC1 Analytical AC1205	Sartryos
pH meter	CG 840	Schott
Photometers	6300	Jen Way
	DU 7400	Beckman
Vortex mixer	7-2020	neoLAB
Rotator	2-1175	neoLAB
Rocking platform	WT16	Biometra
French press	Cell FA-031	SIM-AMINCO
Power supply	Power PAC 300	BIO-RAD
PAGE gel chamber	Min 2-D cell	BIO-RAD
DNA gel chamber	HG370/HG330	Savant
Gel documentation	Video Graphic Printer UR890CE	Sony
	UV Table	Biometra
	Transluminator	Roth
	Video Monitor WV-BM900	Panasonic
	PowerShot A640	Canon
Centrifuges	5417R	Eppendorf
	J2-MC	Beckman
	Biofuge 28RS	Heraeus
Incubator	MIR-153	SANYO
Culture shaker	HT	INFORS
Thermostat bath	F25	Julabo
Thermal block	Thermomixer 5437	Eppendorf
Thermocycler	PTC-200	MJ Research
Chromatography systems	Äkta explorer	GE Healthcare
	LKB-FRAC-200	GE Healthcare
	Uvicord SII	GE Healthcare
	LCC-501 PLUS	GE Healthcare
	LKB-REC-102	GE Healthcare

	LKB-Pump P-500	GE Healthcare
	Vision Workstation	Applied Bioscience
Chromatography columns	HiTrap Heparin Sepharose HP	GE Healthcare
	Mono S HR 5/5	GE Healthcare
	Superdex 75 HiLoad 16/60	GE Healthcare
Concentrators	Amicon Ultra (5K-cutoff)	Millipore
Dialysis	Membrane tubing MWCO 3,500	Spectra/Por
	Membrane closures	Spectra/Por
Titration calorimeter	VP-ITC	MicroCal
Thermostat	ThermoVac	MicroCal
Analytical ultracentrifuge	XL-A	Beckman
Crystallization plate	V4-90S	Victor Recker
	12-well disposable tray	Nelipak Thermoforming
	Glass plates # 2 (22 x 22 mm)	MENZEL GLÄSER
Crystal observation system	Microscope WILD M3C	Leica
	Camera COOLPIX 990	Nikon
Nylon loops	CryoLoops	Hampton Research
Pins	CrystalCap Copper	Hampton Research
Home X-ray source	dtb345	MAR-Research
	Rotating anode RU H2B	Rigaku
	Mirror system	MaxFlux
	Cryo-cooling	Oxford Cryosystems

2.2 Chemicals

Standard chemicals were purchased from Roth Chemikalien, Sigma-Aldrich and Merck.

Agar, granulated	BD Bioscience
Agarose, electrophoresis grade	Invitrogen
Ammonium persulfate (APS)	Serva
Ampicillin sodium sulfate	Roth
Isopropyl-thiogalactoside (IPTG)	AppliChem
Kanamycin sulfate	Roth
N,N,N',N'-tetramethylethylenediamine (TEMED)	Serva
Nucleic acids	Fluka
Polyethyleneimine (PEI), 50% solution in water	Serva
Rotiphorese acrylamide/bisacrylamide (37.5:1, 30%) solution	Roth
Sodium azide	Serva
Tetracycline-hydrochlorin	Boehringer
Tryptone	AppliChem
Yeast extract	MP Biomedicals

Enzymes

Restriction enzymes	*Dpn*I	NEB
Polymerases	*Pfu*	Stratagene
	Turbo *Pfu*	Stratagene
	Pfu Plus!	Roboklon
Nuclease	DNaseI	Roche

Kits

QIAprep Spin Miniprep Kit	Qiagen
QIAGEN Plasmid Maxiprep Kit	Qiagen
QIAquick PCR Purification/Gel Extraction Kit	Qiagen

2.3 Strains and plasmids

Bacterial strains

E. coli BL21(DE3)	Novagen
E. coli DH5α	Invitrogen

Plasmids

RP4 in HB101 cells	Low-copy plasmid with ampicillin, kanamycin, tetracycline resistance (Pansegrau *et al.* 1994)	Erich Lanka
pJM101a-3	*korA* expression plasmid with ampicillin resistance (Balzer *et al.* 1992)	Erich Lanka

Template clones

human CDK4 (1-240)	Kerstin Böhm
human p115 (1-353)	Harald Striegl

2.4 Media and buffers

Culture media

LB (Luria Broth) medium	Tryptone	10 g l^{-1}
	Yeast extract	5 g l^{-1}
	NaCl	10 g l^{-1}
LB agar	LB medium	1 l
	Agar	15 g

Molecular biology

TAE buffer	TRIS/HCl	40 mM
	Sodium acetate	5 mM
	EDTA	1 mM
	Acetic acid	to pH 8
Ethidium bromide solution	TAE buffer	1 l
	Ethidium bromide	10 g
DNA sample buffer 5x	EDTA	100 mM
	Xylencyanol	0.05% (v/v)
	Glycerol	30% (v/v)
	Bromphenol blue	0.05% (w/v)
	in TAE buffer	
DNA molecular weight marker	1 kb ladder (NEB)	0.5 µg $µl^{-1}$
	100 bp ladder (NEB)	0.5 µg $µl^{-1}$
TFB 1	RbCl	100 mM
	$MnCl_2$	50 mM
	Potassium acetate	30 mM
	$CaCl_2$	10 mM
	Glycerol	15% (v/v)
	Acetic acid	to pH 5.8

TFB 2	RbCl	10 mM
	$CaCl_2$	75 mM
	Glycerol	15% (v/v)
	MOPS	10 mM
	NaOH	to pH 6.8

Protein purification

Lysis buffer	TRIS/HCl	20 mM
	Glycerol	10% (v/v)
	NaCl	1 M
	DTT	1 mM
	Pefablock	1 mM
	DNase	10 µg ml^{-1}
		pH 7.5
Dialysis buffer	NaH_2PO_4	20 mM
	Glycerol	5% (v/v)
	DTT	0.1 mM
		pH 7.0
Wash buffer	NaH_2PO_4	20 mM
	DTT	0.1 mM
		pH 7.0
Elution buffer	NaH_2PO_4	20 mM
	NaCl	2 M
	DTT	0.1 mM
		pH 7.0
Gel filtration buffer	TRIS/HCl	20 mM
	NaCl	100 mM
	$MgCl_2$	3 mM
		pH 8.0

SDS PAGE

Electrophoresis buffer	TRIS	25 mM
	Glycine	190 mM
	SDS	0.1% (w/v)
Stacking gel buffer (4x)	TRIS/HCl	0.5 M
	SDS	0.4% (w/v)
		pH 6.8
Stacking gel	30% AA	0.5 ml
	Stacking gel buffer	1 ml
	H_2O	3 ml
	APS 10% (w/v)	75 µl
	TEMED	10 µl
Separation gel buffer (4x)	TRIS/HCl	1.5 M
	SDS	0.4% (w/v)
		pH 8.8
Separation gel 17.5% (w/v)	30% AA	5.8 ml
	Separation gel buffer	2.5 ml
	H_2O	1.7 ml
	APS 10% (w/v)	75 µl
	TEMED	10 µl
SDS sample buffer 2x	SDS	3% (w/v)
	Glycerol	20% (v/v)
	β-mercaptoethanol	3% (v/v)
	Bromphenol blue	0.05% (w/v)
	EDTA	10 mM
Coomassie stain	Coomassie R-250	0.25% (w/v)
	Ethanol	45% (v/v)
	ddH_2O	45% (v/v)
	Acetic acid	10% (v/v)

Destain solution	Ethanol	40% (v/v)
	ddH_2O	50% (v/v)
	Acetic acid	10% (v/v)

Native PAGE

TBE buffer	TRIS/HCl	89 mM
	Boric acid	89 mM
	EDTA	2 mM
		pH 8.3
Separation gel 7.5%	10xTBE	2 ml
	30% AA	5 ml
	H_2O	13 ml
	APS 10% (w/v)	400 µl
	TEMED	30 µl
Sample buffer 2x	Bromphenol blue	0.05% (w/v)
	Glycerol	50%

Crystallization

All solutions used in crystallization trials were prepared following the Hampton Crystal Screen 1 and 2 formulations given in Table 2.1 and Table 2.2 (Jancarik *et al.* 1991), with the addition of 0.02% (w/v) sodium azide.

Screen 1	Salt	Buffer	Precipitant
1	0.02 M $CaCl_2$	0.1 M $NaCH_3CO_2$ pH 4.6	30% v/v MPD
2	-	-	0.4 M K Na tartrate
3	-	-	0.4 M $(NH_4)H_2PO_4$
4	-	0.1 M TRIS/HCl pH 8.5	2.0 M $(NH_4)_2SO_4$
5	0.2 M Na citrate	0.1 M HEPES Na pH 7.5	30% v/v MPD
6	0.2 M $MgCl_2$	0.1 M TRIS/HCl pH 8.5	30% w/v PEG 4,000
7	-	0.1 M Na cacodylate pH 6.5	1.4 M $NaCH_3CO_2$
8	0.2 M Na citrate	0.1 M Na cacodylate pH 6.5	30% v/v 2-Propanol
9	0.2 M $NH_4CH_3CO_2$	0.1 M Na citrate pH 5.6	30% w/v PEG 4,000
10	0.2 M $NH_4CH_3CO_2$	0.1 M $NaCH_3CO_2$ pH 4.6	30% w/v PEG 4,000
11	-	0.1 M Na citrate pH 5.6	1.0 M $(NH_4)H_2PO_4$
12	0.2 M $MgCl_2$	0.1 M HEPES Na pH 7.5	30% v/v 2-Propanol
13	0.2 M Na citrate	0.1 M TRIS/HCl pH 8.5	30% v/v PEG 400
14	0.2 M $CaCl_2$	0.1 M HEPES Na pH 7.5	28% v/v PEG 400
15	0.2 M $(NH_4)_2SO_4$	0.1 M Na cacodylate pH 6.5	30% w/v PEG 8,000
16	-	0.1 M HEPES Na pH 7.5	1.5 M Li_2SO_4
17	0.2 M Li_2SO_4	0.1 M TRIS/HCl pH 8.5	30% w/v PEG 4,000
18	0.2 M $Mg(CH_3CO_2)_2$	0.1 M Na cacodylate pH 6.5	20% w/v PEG 8,000
19	0.2 M $NH_4CH_3CO_2$	0.1 M TRIS/HCl pH 8.5	30% v/v 2-Propanol
20	0.2 M $(NH_4)_2SO_4$	0.1 M $NaCH_3CO_2$ pH 4.6	25% w/v PEG 4,000
21	0.2 M $Mg(CH_3CO_2)_2$	0.1 M Na cacodylate pH 6.5	30% v/v MPD
22	0.2 M $NaCH_3CO_2$	0.1 M TRIS/HCl pH 8.5	30% w/v PEG4,000
23	0.2 M $MgCl_2$	0.1 M HEPES Na pH 7.5	30% v/v PEG 400
24	0.2 M $CaCl_2$	0.1 M $NaCH_3CO_2$ pH 4.6	20% v/v 2-Propanol
25	-	0.1 M Imidazole pH 6.5	1.0 M $NaCH_3CO_2$
26	0.2 M $NH_4CH_3CO_2$	0.1 M Na citrate pH 5.6	30% v/v MPD
27	0.2 M Na citrate	0.1 M HEPES Na pH 7.5	20% v/v 2-Propanol
28	0.2 M $NaCH_3CO_2$	0.1 M Na cacodylate pH 6.5	30% w/v PEG 8,000
29	-	0.1 M HEPES Na pH 7.5	0.8 M K Na tartrate
30	0.2 M $(NH_4)_2SO_4$	-	30% w/v PEG 8,000
31	0.2 M $(NH_4)_2SO_4$	-	30% w/v PEG 4,000
32	-	-	2.0 M $(NH_4)_2SO_4$
33	-	-	4.0 M Na formate
34	-	0.1 M $NaCH_3CO_2$ pH 4.6	2.0 M Na formate
35	-	0.1 M HEPES Na pH 7.5	0.8 M NaH_2PO_4 0.8 M KH_2PO_4
36	-	0.1 M TRIS/HCl pH 8.5	8% w/v PEG 8,000
37	-	0.1 M $NaCH_3CO_2$ pH 4.6	8% w/v PEG 4,000
38	-	0.1 M HEPES Na pH 7.5	1.4 M Na citrate
39	-	0.1 M HEPES Na pH 7.5	2% v/v PEG 400 2.0 M $(NH_4)_2SO_4$
40	-	0.1 M Na citrate pH 5.6	20% v/v 2-Propanol 20% w/v PEG 4,000
41	-	0.1 M HEPES Na pH 7.5	10% v/v 2-Propanol 20% w/v PEG 4,000
42	0.05 M KH_2PO_4	-	20% w/v PEG 8,000
43	-	-	30% w/v PEG 1,500
44	-	-	0.2 M Mg formate
45	0.2 M $Zn(CH_3CO_2)_2$	0.1 M Na cacodylate pH 6.5	18% w/v PEG 8,000
46	0.2 M $Ca(CH_3CO_2)_2$	0.1 M Na cacodylate pH 6.5	18% w/v PEG 8,000
47	-	0.1 M $NaCH_3CO_2$ pH 4.6	2.0 M $(NH_4)_2SO_4$
48	-	0.1 M TRIS/HCl pH 8.5	2.0 M $(NH_4)H_2PO_4$
49	1.0 M Li_2SO_4	-	2% w/v PEG 8,000
50	0.5 M Li_2SO_4	-	15% w/v PEG 8,000

Table 2.1: Hampton Crystal Screen 1 formulations.

Screen 2	Salt	Buffer	Precipitant
1	2.0 M NaCl	-	10% w/v PEG 6,000
2	0.5 M NaCl 0.01 M $MgCl_2$	-	0.01 M CTAB
3	-	-	25% v/v Ethylene glycol
4	-	-	35% v/v 1,4-Dioxane
5	2.0 M $(NH_4)_2SO_4$	-	5% v/v 2-Propanol
6	-	-	1.0 M Imidazole pH 7.0
7	-	-	10% w/v PEG 1,000 10% w/v PEG 8,000
8	1.5 M NaCl	-	10% v/v Ethanol
9	-	0.1 M $NaCH_3CO_2$ pH 4.6	2.0 M NaCl
10	0.2 M NaCl	0.1 M $NaCH_3CO_2$ pH 4.6	30% v/v MPD
11	0.01 M $CoCl_2$	0.1 M $NaCH_3CO_2$ pH 4.6	1.0 M 1,6-Hexanediol
12	0.1 M $CdCl_2$	0.1 M $NaCH_3CO_2$ pH 4.6	30% v/v PEG 400
13	0.2 M $(NH_4)_2SO_4$	0.1 M $NaCH_3CO_2$ pH 4.6	30% w/v PEG MME 2,000
14	0.2 M K Na tartrate	0.1 M Na citrate pH 5.6	2.0 M $(NH_4)_2SO_4$
15	0.5 M $(NH_4)_2SO_4$	0.1 M Na citrate pH 5.6	1.0 M Li_2SO_4
16	0.5 M NaCl	0.1 M Na citrate pH 5.6	2% v/v PEI
17	-	0.1 M Na citrate pH 5.6	35% v/v tert-Butanol
18	0.01 M $FeCl_3$	0.1 M Na citrate pH 5.6	10% v/v Jeffamine M-600
19	-	0.1 M Na citrate pH 5.6	2.5 M 1,6-Hexanediol
20	-	0.1 M MES pH 6.5	1.6 M $MgSO_4$
21	0.1 M NaH_2PO_4 0.1 M KH_2PO_4	0.1 M MES pH 6.5	2.0 M NaCl
22	-	0.1 M MES pH 6.5	12% w/v PEG 20,000
23	1.6 M $(NH_4)_2SO_4$	0.1 M MES pH 6.5	10% v/v 1,4-Dioxane
24	0.05 M CsCl	0.1 M MES pH 6.5	30% v/v Jeffamine M-600
25	0.01 M $CoCl_2$	0.1 M MES pH 6.5	1.8 M $(NH_4)_2SO_4$
26	0.2 M $(NH_4)_2SO_4$	0.1 M MES pH 6.5	30% w/v PEG MME 5,000
27	0.01 M $ZnSO_4$	0.1 M MES pH 6.5	25% v/v PEG MME 550
28	-	-	1.6 M Na citrate pH 6.5
29	0.5 M $(NH_4)_2SO_4$	0.1 M HEPES Na pH 7.5	30% v/v MPD
30	-	0.1 M HEPES Na pH 7.5	10% w/v PEG 6,000 5% v/v MPD
31	-	0.1 M HEPES Na pH 7.5	20% v/v Jeffamine M-600
32	0.1 M NaCl	0.1 M HEPES Na pH 7.5	1.6 M $(NH_4)_2SO_4$
33	-	0.1 M HEPES Na pH 7.5	2.0 M NH_4 formate
34	0.05 M $CdSO_4$	0.1 M HEPES Na pH 7.5	1.0 M $NaCH_3CO_2$
35	-	0.1 M HEPES Na pH 7.5	70% v/v MPD
36	-	0.1 M HEPES Na pH 7.5	4.3 M NaCl
37	-	0.1 M HEPES Na pH 7.5	10% w/v PEG 8,000 8% v/v Ethylene glycol
38	-	0.1 M HEPES Na pH 7.5	20% w/v PEG 10,000
39	0.2 M $MgCl_2$	0.1 M TRIS/HCl pH 8.5	3.4 M 1,6-Hexanediol
40	-	0.1 M TRIS/HCl pH 8.5	25% v/v tert-Butanol
41	0.01 M $NiCl_2$	0.1 M TRIS/HCl pH 8.5	1.0 M Li_2SO_4
42	1.5 M $(NH_4)_2SO_4$	0.1 M TRIS/HCl pH 8.5	12% v/v Glycerol
43	0.2 M $(NH_4)H_2PO_4$	0.1 M TRIS/HCl pH 8.5	50% v/v MPD
44	-	0.1 M TRIS/HCl pH 8.5	20% v/v Ethanol
45	0.01 M $NiCl_2$	0.1 M TRIS/HCl pH 8.5	20% w/v PEG MME 2,000
46	0.1 M NaCl	0.1 M BICINE pH 9.0	20% v/v PEG MME 550
47	-	0.1 M BICINE pH 9.0	2.0 M $MgCl_2$
48	-	0.1 M BICINE pH 9.0	2% v/v 1,4-Dioxane 10% w/v PEG 20,000

Table 2.2: Hampton Crystal Screen 2 formulations.

2.5 Synthetic oligonucleotides

2.5.1 Primers

Primers were purchased from Eurofins MWG Operon, Ebersberg. Forward primers are labeled with cw (clockwise) and reverse primers with ccw (counterclockwise).

primer	Sequences (5' → 3')	Restriction site
KA_cwB001[a,b]	aat gtt agt taa gga tcc atg aag aaa cgg ctt acc gaa	*Bam*HI
KA_ccwH101[a,b]	aat gtt agt taa aag ctt tca tcg ttt ggt ttc ctg ttt t	*Hind*III
O_A1_cw[b]	aat gtt agt taa cag ctg tgt ggc ttc cca tcg act aa	*Pvu*II
O_A1_ccw[b]	aat gtt agt taa aag ctt atc ggc gcg atc ctg gcg	*Hind*III
TA348_cw[b]	aat gtt agt taa gga tcc atg acg aaa cat gag ctg tcg	*Bam*HI
TA348_ccw[b]	aat gtt agt taa aag ctt tca gag cct tcc acg cag c	*Hind*III
K745_cw[c]	tag att agg atc cat ggc tac ctc tcg ata tga g	*Bam*HI
K745_ccw[c]	aat gtt agt tgc ggc cgc tca cag gga tac atc tcg agg cc	*Not*I
H1082_cw[d]	gcg gat cca att tcc tcc gcg ggg	*Bam*HI
H1082_ccw[d]	Gtt agc ggc cgc tca aac ccc agt agc cat ta	*Not*I

[a]Primer for the pJM101a-3 plasmid.
[b]Primer for the RP4 plasmid.
[c]Primer for the template clone human CDK4 (1-240).
[d]Primer for the template clone human p115 (1-353).

2.5.2 Oligonucleotides for QuikChange mutagenesis

Only forward primers are listed. Reverse primers have the reverse complementary sequence. Oligonucleotides were purchased from Eurofins MWG Operon, Ebersberg.

primer	**Sequences (5' → 3')**
KA_R48A_cw	caa cgt cgc tgg gac tga ccg cgg gcg cag tgt cgc aag cgg t
KA_Q53A_cw	tga cca ggg gcg cag tgt cgg cag cgg tgc atc gcg tgt ggg c
KA_Q53E_cw	tga cca ggg gcg cag tgt cgg aag cgg tgc atc gcg tgt ggg c
KA_QR48A_cw	caa cgt cgc tgg gac tga ccg cgg gcg cag tgt cgg cag cgg t

2.5.3 Oligonucleotides for ITC and crystallization

Oligonucleotides were purchased from BioTeZ, Berlin.

Oligonucleotide	**Sequence (5' → 3')**
O_A strand 1	CTT GTT TAG CTA AAG ATT
O_A strand 2	AAT GTT TAG CTA AAC AAG
O_A* strand 1	$C^{Br}U^{Br}U$ GTT TAG CTA AAC $A^{Br}UT$
O_A* strand 2	AAT GTT TAG CTA AAC AAG

3 Methods

3.1 Molecular biology methods

3.1.1 Polymerase chain reaction (PCR)

PCR (Mullis and Faloona 1987) is a technique that allows the amplification of a specific fragment of double-stranded DNA. In this study, it was used to obtain oligonucleotides for EMSA experiments (3.2.11). First, the template DNA (2.3) that contains the target DNA to be amplified is heated to denature (melt) the double-stranded DNA duplex. Second, the solution is cooled in the presence of an excess of two 20 – 40 bp long synthetic single-stranded oligonucleotides (primers, see 2.5.1) that are complementary to the DNA sequence flanking the target DNA. DNA synthesis always occurs in the 5' to 3' direction (reading the template strand 3' to 5'). Therefore, the primers must be complementary to opposite strands of the DNA duplex. Third, a heat-stable DNA polymerase is added, along with the four deoxyribonucleotide triphosphates (dNTPs), so that two new DNA strands can be synthesized by primer extension. If this melting, annealing, and polymerization cycle is repeated, the fragment of double-stranded DNA located between the primer sequences can be amplified over a millionfold in a matter of hours.

The reagents used for a PCR setup are listed in Table 3.1. The heat-stable DNA polymerase (*Pfu*) has proofreading activity to reduce the probability of introducing point mutations (single base changes) in the amplified DNA product. In Table 3.2 the standard PCR program is given. PCR reactions were purified using the QIAquick PCR purification kit from Qiagen according to the manufacturer.

Compound	**Concentration**
H_2O	to 50 µl
template DNA	10 – 15 ng
forward primer	0.6 µM
reverse primer	0.6 µM
dNTPs each	250 µM
10x *Pfu* reaction buffer	1x
DMSO	6.25 % (v/v)
Pfu polymerase	5 U

Table 3.1: Standard PCR setup.

Segment	Task	Temperature [°C]	Time [min:sec]
1. Initialization	melting of DNA	95	1:00
2. 25 cycles	melting of DNA	95	0:30
	primer annealing	55	1:00
	primer extension	68	2:00
3.	final extension	68	10:00
4. Termination	hold	4	forever

Table 3.2: Standard PCR program.

3.1.2 Gel electrophoresis

Gel electrophoresis was used to determine the size and to estimate the amount of PCR products for the EMSA experiments (3.2.11). 0.3 g agarose was mixed with 30 ml TAE buffer and heated in a microwave until the agarose had completely dissolved. The 1% agarose gel was filled into a gel chamber and supplemented with 0.5 µg ml^{-1} ethidium bromide. Samples were mixed with DNA sample buffer and loaded on the gel, as well as a DNA ladder with known band sizes and concentrations for estimation purposes. Electrophoresis was performed at 220 mA in TAE buffer, and the gel was analyzed under UV light.

3.1.3 Annealing of DNA oligonucleotides

Single stranded HPLC-purified oligonucleotides (2.5.3) were synthesized by BioTeZ, Berlin. Equimolar amounts of complementary strands were solubilized and mixed in gel filtration buffer (2.4). The solution was heated to 94 °C in a thermostat water bath, which was then allowed to cool to room temperature overnight. The annealed DNA duplexes were further purified on a gel filtration column Superdex 75, equilibrated with gel filtration buffer.

3.1.4 Determination of DNA concentration

To determine the concentration of DNA preparations a UV-vis spectrometer was used recording spectra in the range of 200 to 350 nm. Buffer without DNA was used as reference. Light absorption at 260 nm was used to calculate the DNA concentration according to Lambert-Beer's law (equation 3.1). The molar extinction coefficients (ε, see Table 3.3) were calculated with the program *OligoCalc* (Kibbe 2007).

$$C = \frac{A_\lambda}{\varepsilon_\lambda l} \qquad (3.1)$$

A_λ : absorbance at wavelength λ
ε_λ : molar extinction coefficient [M^{-1} cm^{-1}] at wavelength λ
C : molar concentration [M]
l : optical path length [cm]

ds DNA	**ε_{260} [M^{-1} cm^{-1}]**
O_A, O_A*	230,415
O_A1	5,586,592
TA348	4,464,286
K745	9,523,810
H1082	13,888,889

Table 3.3: Molar extinction coefficients derived from *OligoCalc* (Kibbe 2007).

3.1.5 QuikChange mutagenesis (QCM)

Single mutants of KorA were generated following a two-stage QuikChange Mutagenesis protocol (Wang and Malcolm 1999). Two complementary 43-bp oligonucleotides (2.5.2) were designed for the template plasmid pJM101a-3 (2.3), containing the desired mutation directly in the center of their sequence. To prevent formation of primer dimers, the first steps of mutagenesis were performed with both primers in separate reaction setups (Table 3.4). After five PCR cycles both reactions were mixed, and additional PCR cycles were performed. The QuikChange mutagenesis program is listed in Table 3.5.

Compound	**Reaction I**	**Reaction II**
H_2O	to 50 µl	to 50 µl
template DNA	40 - 100 ng	40 - 100 ng
forward primer	0.5 µM	--
reverse primer	--	0.5 µM
dNTPs each	400 µM	400 µM
10x *Pfu* reaction buffer	1x	1x
Turbo *Pfu* or *Pfu* Plus! polymerase	5 U	5 U

Table 3.4: Standard QCM reaction setup.

The amplification procedure is followed by incubation of the DNA with *Dpn*I restriction enzyme overnight at 37 °C. *Dpn*I cleaves the methylated original template plasmid which was isolated from *E. coli* cells. The PCR-generated synthetic DNA remained intact, and was transformed into chemically competent *E. coli* DH5α cells (3.1.7). Plasmids were selected, isolated (3.1.6) and checked by sequencing.

Segment	**Task**	**Temperature [°C]**	**Time [min:sec]**
1. Initialization	melting of DNA	95	1:00
2. 5 cycles	melting of DNA	95	0:30
	primer annealing	55	1:00
	primer extension	68	10:00
3. Pause		reactions I and II were mixed 1:1	
4. Initialization	melting of DNA	95	1:00
5. 25 cycles	melting of DNA	95	0:30
	primer annealing	55	1:00
	primer extension	68	10:00
6. Termination	hold	4	forever

Table 3.5: Standard QCM program.

3.1.6 Plasmid preparations

Colonies picked from agar plates were grown overnight in LB medium for mini (4 ml) or maxi (150 ml) plasmid preparations. The plasmids were isolated using the QIAprep Spin Miniprep Kit and QIAGEN Plasmid Maxiprep Kit according to the manufacturer.

3.1.7 Transformation of chemically competent *E. coli* cells

Artificial transformation of *E. coli* cells was performed with chemically competent cells according to the protocol by Hanahan (1983). 1 ml LB medium was inoculated with a sample of *E. coli* BL21(DE3) or *E. coli* DH5α (2.3) obtained from a stock solution and grown overnight. The culture was diluted 1:100 in fresh LB medium and grown at 37 °C to A_{600} = 0.6. Cells were cooled down on ice and harvested at 4 °C by centrifugation for 5 min at 2,500x *g*. The pellet was resuspended in 30 ml prechilled TFB 1 buffer (2.4), incubated for 45 min on ice and centrifuged for 5 min at 2,500x *g*. Cells were then resuspended in 8 ml prechilled TFB 2 buffer (2.4), and 50 µl aliquots were flash-frozen in liquid nitrogen and stored at -70 °C.

Competent cell cultures were thawed on ice before DNA was added for transformation. Reactions were incubated on ice for 30 min, followed by a heat shock at 42 °C for 42 sec. 800 µl LB medium was added immediately, and cells were shaken at 37 °C. After 50 min the cultures were streaked out on LB plates containing appropriate antibiotics.

3.2 Protein purification and characterization methods

3.2.1 Expression

E. coli BL21(DE3) cells carrying the appropriate wild-type or mutant *korA* expression plasmid were cultured at 37 °C in LB medium containing 100 µg ml^{-1} of ampicillin. An overnight pre-culture was diluted 50-fold in the same medium and grown to A_{600} = 0.8. Transcription in cells was induced by addition of 1 mM IPTG, and growth was continued for 4 h. Cells were harvested by centrifugation at 6,000x *g* for 8 min and pellets were stored at -80 °C.

3.2.2 Cell lysis

The cell pellets were resuspended in lysis buffer (2.4). During lysis the cells were constantly cooled. The suspension was passed through the French Press twice at 1100 psi (7.6 MPa). The lysate was cleared by centrifugation at 45,000x *g* for 60 min.

3.2.3 Precipitation with polyethyleneimine

Nucleic acids were removed from the lysate, by precipitation with polyethyleneimine (PEI) at 0.25%. 50% PEI solution is very viscous and was therefore diluted 10-fold in lysis buffer (2.4). The pH of PEI containing lysis buffer was adjusted to 7.5 before adding it to the lysate. The suspension was stirred for 20 min on ice, and the precipitate was removed by centrifugation at 45,000x *g*.

3.2.4 Ammonium sulfate precipitation (ASP)

Ammonium sulfate is the most commonly used salt to fractionate proteins by precipitation. When neutral salts are added to the protein lysate, the salt ions are solvated by water molecules. Increasing salt concentration removes the water molecules from around the protein, exposing hydrophobic patches at their surface. Proteins aggregate and precipitate in fractions due to the difference in amount and size of their hydrophobic patches.

Ammonium sulfate was added to the lysate, and the pH was maintained between 7.0 - 7.5. The fraction of proteins which precipitated between 30 – 50% of ammonium sulfate saturation was enriched in KorA protein. The suspension was stirred for 20 min on

ice, before the precipitate was removed by centrifugation at 45,000x *g*. The pellet was dissolved in dialysis buffer (2.4), using twice the volume of the pellet.

3.2.5 Dialysis

It is necessary to reduce high salt concentrations (e.g. ammonium sulfate, sodium chloride) from the buffer before working with chromatography steps where the ionic strength is increased to desorb proteins from the ligands attached to the column matrix (see 3.2.6 and 3.2.7).

This can be achieved by dialysis, where the protein solution is placed inside a semipermeable membrane tubing which is immersed in 1 – 2 l of stirred dialysis buffer (2.4). For the removal of ammonium sulfate the dialysis buffer was changed twice within the first half of the dialysis (6 h). Sodium chloride was removed by dialyzing for 3 h, with one change of buffer. For isothermal titration calorimetry (3.2.13) purified KorA protein and O_A oligonucleotides were dialyzed against the gel filtration buffer (2.4) overnight.

3.2.6 Affinity chromatography

Affinity chromatography is highly selective, and if the affinity of the column ligand for the protein of interest is sufficiently high, it allows purification of a single protein from a crude mixture of proteins with a large volume. The HiTrap Heparin Sepharose HP column binds proteins which contain a DNA-binding motif or exhibit a positive net charge. The porcine heparin ligand is coupled to the Sepharose High Performance base matrix *via* the N-hydroxysuccinamide coupling method. Heparin, a sulfated glycosaminoglycan, has a strong negative charge. Elution is achieved by addition of a high-salt buffer which will occupy the binding sites of heparin if present in excess.

Nucleic acids were removed from the lysate by PEI precipitation (3.2.3) before using the HiTrap Heparin column. After binding, the affinity matrix was washed extensively with six column volumes of wash buffer (2.4), before eluting the bound proteins by running a linear gradient of 0 – 50% of elution buffer (2.4). All chromatography experiments (3.2.6 – 3.2.8) were performed with Pharmacia FPLC systems or the Äkta Explorer.

3.2.7 Cation-exchange chromatography

In ion-exchange chromatography proteins are separated on the basis of their surface charge. Separation of proteins is achieved by their difference in equilibrium distribution between a buffered mobile phase and a stationary phase consisting of a matrix to which charged inorganic groups are attached. Cation-exchange matrices are derivatized with negatively-charged groups for the adsorption of cationic proteins. The ionogenic groups used in Mono S HR 5/5 columns are strong sulfo groups. The net charge on a protein varies with the pH of the buffer. At pH values below their pI proteins carry an overall positive charge.

After adsorption of the proteins to the charged groups on the matrix, the Mono S column was washed with five column volumes of wash buffer (2.4), before eluting the bound proteins by running a linear gradient of 0 – 50% of elution buffer (2.4).

3.2.8 Gel-filtration chromatography

Gel filtration is a partition chromatography that separates molecules based on their mass and shape. Particles are partitioned between solvent and a stationary phase of defined porosity. Smaller molecules can enter the matrix pores and therefore move slower through the column, while larger molecules do not fit inside most pores and elute first from the column. Separation in the order of decreasing sizes is not possible when molecules are not of globular shape. Longer stretched particles travel shorter ways through the column than globular molecules of the same size. The pore size of the Superdex 75 HiLoad 16/60 allows separation of proteins of about 3 - 70 kDa. Larger proteins are excluded from the stationary phase. Their elution volume is the column specific void volume.

Each gel filtration column has its own characteristics and therefore needs to be calibrated prior to the experiment. The elution volumes of suitable standard proteins are measured and plotted against the logarithm of the molecular weight. To determine the size of a molecule of interest, its elution volume is used to obtain the molecular weight from the calibration graph.

The column was equilibrated with two column volumes of gel filtration buffer (2.4). The maximal sample volume was 2% of the column volume. The elution profile was monitored measuring the absorption at 280 nm, 260 nm, and 220 nm to separate between protein, DNA and protein-DNA complexes.

3.2.9 Determination of protein concentration

Light absorption at 280 nm was used to calculate the protein concentration according to Lambert-Beer's law (see 3.1.4). The molar extinction coefficient (ε) of a protein was calculated with *ProtParam* (Gasteiger *et al.* 2005).

KorA proteins	ε_{280} [M^{-1} cm^{-1}]	A_{280} at 1 mg ml^{-1}
wt	13,980	1.237
R48A	13,980	1.246
Q53A	13,980	1.243
R48A/Q53A	13,980	1.252
Q53E	13,980	1.236

Table 3.6: Molar extinction coefficients calculated with *ProtParam* (Gasteiger *et al.* 2005).

3.2.10 SDS and native PAGE

Polyacrylamide gels are used in gel electrophoretic separation of a wide array of proteins and nucleic acids. The gels are formed as polymerization of acrylamide and the crosslinker N-N'-methylene-bis-acrylamide is induced in the presence of a free-radical initiator ammonium persulfate (APS) and the catalyst N,N,N',N'-tetramethylethylene-diamine (TEMED). Variations in the concentration of both monomers modificate the crosslinking and thus the pore size of the gel. The pore size required for a particular electrophoretic separation will depend on the difference in size of the protein or DNA samples to be resolved.

Sodium dodecyl sulfate (SDS) is an anionic detergent that denatures proteins. It binds to all regions of the proteins and disrupts most non-covalent intermolecular and intramolecular protein interactions. The intrinsic charge on the individual polypeptide chains is masked and the charge-to-mass ratio of all proteins in the sample is approximately the same. The migration velocity through the gel is proportional to the logarithm of their molecular mass.

Prior to SDS PAGE, samples were mixed with SDS sample buffer (2.4) and injected into wells within the stacking gel. Electrophoresis was carried out with electrophoresis buffer (2.4) at 130 V for migration through the stacking gel and at 180 V for migration through the separation gel. Proteins are focused during migration through the stacking gel and enter

the separation gel all at once. The comparison to a marker lane with proteins of known size allows an estimation of the relative molecular mass of the sample.

For native polyacrylamide gel electrophoresis, gels were prepared according to the standard Laemmli SDS-PAGE protocol (Laemmli 1970) for the stacking gel without adding a denaturing (SDS) or reducing (DTT) detergent. Proteins retain their native structure and also their functions. Migration velocity through the gel now depends on many factors, for example, size, shape, and intrinsic charge.

3.2.11 Electrophoretic mobility shift assay (EMSA)

EMSA was used for identifying protein-DNA interactions. From a native polyacrylamide gel the binding affinity and specificity of a DNA-binding protein to oligonucleotides can be obtained. DNA duplexes that are bound by protein migrate more slowly compared to their unbound state. Bands of these DNA fragments are shifted relative to the bands from the unbound duplexes. A 436-bp DNA fragment containing O_A1 (Table 1.2) was PCR-amplified using the RP4 plasmid (2.3 and 2.5.1) as template. The competitor DNA of 348 bp, 745 bp and 1082 bp in length, without the O_A sequence, was also generated by PCR (see 3.1.1) using RP4, and the template clones for human CDK4 (1-240) and human p115 (1-353) (see 2.3 and 2.5.1), respectively. The standard reaction mixture (10 µl) contained gel filtration buffer (2.4) and 0.15 pmol of each PCR construct. The reactions were then incubated with purified protein for 20 min at 37 °C. After incubation the reactions were analyzed by electrophoresis on a non-denaturing 7.5% polyacrylamide gel in TBE buffer at 8 V cm^{-1}, stained with ethidium bromide.

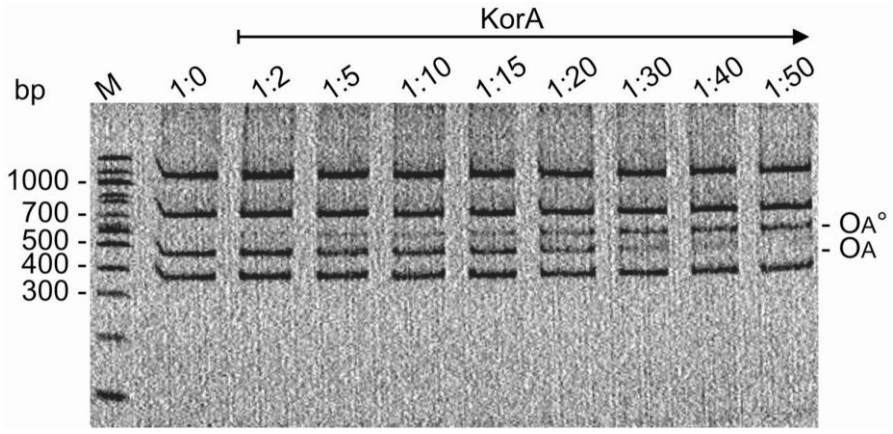


Figure 3.1: Testing for DNA binding of KorA. The 436-bp fragment carries the class I O_A site (*korAp*) of plasmid RP4, while the other fragments serve as competitor DNA. Increasing amounts of wt KorA were incubated with a DNA mixture containing 0.15 pmol of each fragment. O_A, DNA fragment containing O_A; O_A°, complex of O_A with wt KorA; M, 100-bp DNA ladder.

3.2.12 Gel staining procedures

For staining with Coomassie Brilliant Blue, gels were boiled 30 s in Coomassie stain solution (2.4). The gel was rinsed with water, then boiled for 30 s in destain solution (2.4) and shaken at room temperature until it was completely destained.

Silver staining was done according to a 30-min protocol shown in Table 3.7 (Nesterenko *et al.* 1994).

Steps	**Solution**	**Time**
Fixation	60 ml acetone stock[a]; 1.5 ml TCA stock[a]; 25 µl 37% HCHO	5 min
Rinse	H_2O	3 × 5 s
Wash	H_2O	5 min
Rinse	H_2O	3 × 5 s
Pretreat	60 ml acetone stock[a]	5 min
Pretreat	100 µl $Na_2S_2O_3$ stock[a]	1 min
Rinse	H_2O	3 × 5 s
Impregnate	0.8 ml $AgNO_3$ stock[a]; 0.6 ml 37% HCHO; 60 ml H_2O	8 min
Rinse	H_2O	3 × 5 s
Development	1.2 g Na_2CO_3; 25 µl HCHO; 25 µl $Na_2S_2O_3$ stock[a]; 60 ml H_2O	10 – 20 s
Stop	1% glacial acetic acid in 60 ml H_2O	30 s
Rinse	H_2O	10 s

[a]Solutions: 50% acetone in H_2O; 50% TCA in H_2O; 20% $AgNO_3$ in H_2O (store in dark); 10% $Na_2S_2O_3$ in H_2O.

Table 3.7: Silver staining protocol.

3.2.13 Isothermal titration calorimetry (ITC)

ITC allows to directly measure the heat change during complex formation at a constant temperature. One binding partner is titrated stepwise into a solution containing the interaction partner, while the calorimeter observes and quantifies the released or absorbed heat deriving from exothermic or endothermic reactions, respectively. A standard ITC graph consists of two panels, as demonstrated in Figure 3.2. The upper panel shows the heat trace of the thermostat with the individual injections of the ligand as peaks. The area of the peaks is integrated and plotted against the molar ratio of ligand and protein, resulting in the data points depicted in the lower panel. A quadratic binding curve can be fitted into the data, called the binding isotherm. The shape of this curve allows simultaneous determination of the binding constant (K), the enthalpy (ΔH°), the entropy (ΔS°), and the stoichiometry (N) (Wiseman *et al.* 1989).

For a reversible reaction of protein P and ligand L the association constant K_A is described by:

$$P + L \leftrightarrow LP \quad \rightarrow \quad K_A = \frac{[LP]}{[P] \cdot [L]} \tag{3.2}$$

In the equilibrium state the total concentration of ligand or protein is:

$$[L]_{tot} = [L] + [LP] \quad \rightarrow \quad [L] = [L]_{tot} - [LP] \tag{3.3}$$

$$[P]_{tot} = [LP] + [P] \quad \text{with equation 3.2:} \quad [P]_{tot} = [LP] + \frac{[LP]}{K_A \cdot [L]} \tag{3.4}$$

equation 3.3 in 3.4:
$$[P]_{tot} = [LP] + \frac{[LP]}{K_A \cdot ([L]_{tot} - [LP])} \tag{3.5}$$

Transformation of equation 3.5 results in a quadratic equation of [LP]:

$$ax^2 + bx + c = 0 = [LP]^2 + [LP] \cdot \left(-[P]_{tot} - [L]_{tot} - \frac{1}{K_A} \right) + [P]_{tot} \cdot [L]_{tot}$$

$\rightarrow x = [LP]$, $a = 1$, $b = -[P]_{tot} - [L]_{tot} - \frac{1}{K_A}$, and $c = [P]_{tot} \cdot [L]_{tot}$

The solution is:
$$[LP] = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{3.6}$$

The derivative of [LP] over $[L]_{tot}$ of equation 3.6 is the change in ligand-protein concentration at increasing ligand concentration:

$$dQ = \frac{d[LP]}{d[L]_{tot}} = \frac{1}{2} + \frac{1 - \frac{1+u}{2} - \frac{w}{2}}{\sqrt{w^2 - 2 \cdot w \cdot (1-u) + (1+u)^2}} \tag{3.7}$$

$$\text{with} \quad u = \frac{1}{K_A \cdot [P]_{tot}} \tag{3.8}$$

$$\text{and} \quad w = \frac{[L]_{tot}}{[P]_{tot}} \tag{3.9}$$

The heat change is proportional to the formation of protein-ligand complexes, with the constants of the cell volume and the molar reaction enthalpy ΔH^0:

$$dQ = V \cdot \Delta H^0 \cdot d[LP] \tag{3.10}$$

equation 3.10 in 3.7:

$$\frac{dQ}{V \cdot d[L]_{tot}} = \Delta H^0 \left(\frac{1}{2} + \frac{1 - \frac{1+u}{2} - \frac{w}{2}}{\sqrt{w^2 - 2 \cdot w \cdot (1-u) + (1+u)^2}} \right) \tag{3.11}$$

$\frac{dQ}{d[L]_{tot}}$ is the parameter which is obtained during the experiment.

Equations 3.8, 3.9 and 3.11 show the coherence of the titration curve to the association constant, concentration of protein and ligand, as well as the molar enthalpy.

If K_A is very large (∞), u (equation 3.8) will be zero and the measured heat depends only on the injected amount of ligand (see equation 3.9). All ligand will be bound as long as unbound protein is available. At the equivalent point, binding does not occur, and thus no heat change is detected.

If the K_A value is very small, the titration curve is a nearly linear function. There is no heat change due to the absence of ligand binding.

If the association constant is in the intermediate range, the shape of the titration curve is sigmoidal. The S shape depends on the ratio between association constant and protein concentration, described by a parameter c, which is reciprocal to u.

$$c = \frac{1}{u} = K_A \cdot [P]_{tot} \tag{3.12}$$

Titration curves in the range of $10 < c < 100$ can be fitted well, see Figure 3.2. The inflection point yields the stoichiometry N of the reaction, and the difference in height of the two plateaus gives the reaction enthalpy ΔH. The association constant is obtained from the curve fit and equation 3.11. Finally, the Gibbs standard enthalpy ΔG^0 and the standard entropy ΔS^0 can be obtained with equation 3.13.

$$\Delta G^0 = -RT \ln K_A = RT \ln K_D = \Delta H^0 - T\Delta S^0 \tag{3.13}$$

For ITC experiments purified KorA and 18-bp oligonucleotides (O_A, see 2.5.3) were dialyzed against gel filtration buffer (2.4). Samples were degassed for ~ 5 min by vacuum aspiration prior to loading, and all titrations were carried out at 310 K. Experiments were performed on a VP-ITC titration calorimeter using the *VPViewer 2000* software for instrument control and data acquisition. 40 µM KorA solution was filled in a stirred (310 rpm) reaction cell of 1.4 ml. Injections, each of 10 µl volume (first injection always 5 µl volume) and 10 s duration with a 4 min interval between injections, were carried out using a syringe filled with 200 µM 18mer O_A solution. Thermogram analysis was performed using the software *ORIGIN* (version 7.0) provided by the manufacturer.

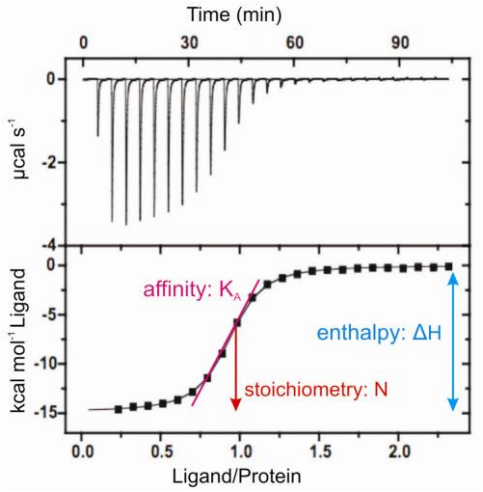


Figure 3.2: ITC thermogram analysis. The upper panel shows the heat trace of the thermostat for an exothermic reaction with individual injections (peaks). From the binding isotherm in the lower panel the parameters for affinity, stoichiometry, and enthalpy can be obtained directly. The graph was adapted from: www.endocytosis.org.

3.2.14 Analytical ultracentrifugation

Analytical equilibrium ultracentrifugation was used to determine the molecular mass of KorA protein in solution by sedimentation, and thus its multimeric state. The motion of particles is influenced by Brown's diffusion, and the applied acceleration of the system must be very strong for sedimentation to occur. The sedimentation rate depends on various parameters, for example, the molecular mass of the particle, the applied angular velocity, the solvent and the temperature.

6-channel cells containing 70 µl of several different sample concentrations in the range of 0.15 – 0.88 mg ml^{-1} were centrifuged against gel filtration buffer (2.1.4) for 2 h at 24,000 rpm followed by 20,000 rpm to equilibrium at 10 °C. The radial concentration distributions of each sample in sedimentation equilibrium were recorded at three different wavelengths between 220 and 270 nm (usually 280, 285, and 290) and fitted to equation 3.14 using the program *POLYMOLE* (Behlke *et al.* 1997). The obtained mass of KorA at different concentrations was extrapolated to infinite dilution.

$$A_r = A_{rm} \cdot e^{MF} \tag{3.14}$$

A_r : radial absorbance at position r
A_{rm} : radial absorbance at meniscus position
M : molecular weight

with $$F = \frac{\left(1 - \rho\bar{v}\right) \cdot \omega^2 \left(r^2 - r_m^2\right)}{2RT} \tag{3.15}$$

ρ : solvent density
v : partial specific volume of the particle
ω : angular velocity
r_m : position of meniscus
R : gas constant
T : absolute temperature

3.3 Structural methods

X-ray crystallography is a technique that allows us to obtain the order of atoms in a single crystal by irradiation with X-rays. A three-dimensional image of the electron density can be derived from the measured angles and intensities of the scattered beams giving the mean position of each atom. Hence the structure of all materials that form well diffracting crystals can be revealed, such as salts, metals, and even biological macromolecules. Besides nuclear magnetic resonance (NMR), X-ray crystallography is the most generally applied method to determine structure and function of proteins and DNA molecules.

3.3.1 Crystallization

To determine the structure of the macromolecule under study a suitable crystal has to be grown. The crystal should be adequate in size, pure in composition, regular in structure without cracks or twinning deficiencies, and of good diffraction quality. It is difficult to arrange protein molecules in a periodically packed crystal due to their irregular shapes and conformational flexibility. Therefore, these crystals are in general very fragile and soft.

Crystals are grown from aqueous solutions of purified protein at high concentration. The solubility of the protein is gradually reduced by increasing precipitant and protein concentration. In a perfect setup, nuclei formation sets in, followed by growth of the crystallites. However, if the protein concentration is too high, many nuclei are formed which cannot grow any further, and if it is too low, growth of crystals is reduced or nucleation is impeded.

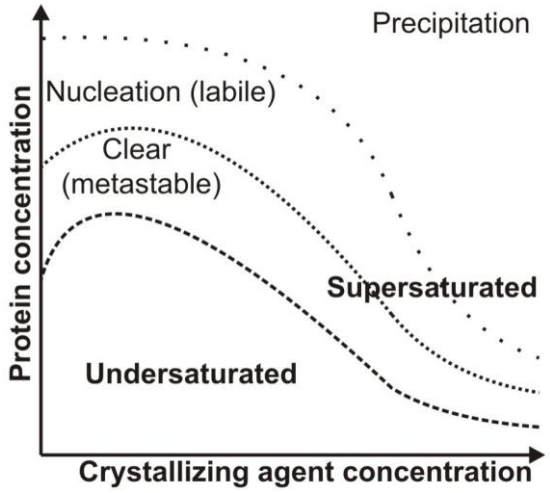


Figure 3.3: Protein crystallization phase diagram. Adapted from McPherson (1999).

The phase diagram in Figure 3.3 illustrates the four zones of protein crystallization that depend on protein saturation and precipitant concentration. Proteins precipitate at high supersaturation, nuclei may form at moderate supersaturation, crystals are stable and may grow at lower supersaturation while nucleation is not possible, and the drop will stay clear when the protein solution is undersaturated. Transition between two phases can take place when, for example, nuclei have formed and the protein concentration therefore drops to a degree optimal for crystal growth. Hence, the process of crystallization should start with higher protein saturation leading first to nucleation and later to metastable conditions.

During crystallization the tertiary structure of the protein must be stably maintained. Therefore, the optimal crystallization solution favoring the growth of a single crystal, and not that of disordered aggregates, needs to be identified by screening many different conditions. Many factors have to be considered to lower the solubility, such as concentration and purity of the protein, precipitants, pH, salt content, temperature, and different additives. Often used as precipitants are ammonium sulfate, PEGs 4000 to 8000 and MPD. The pH can be adjusted using different buffers that affect the surface charge on the protein and therefore its interactions with the solvent. Protein oils can be dissolved or avoided by small polar organic additives such as glycerol and 1,6-hexanediol. Initial screening of crystallization conditions is generally done according to the sparse matrix design by Jancarik and Kim, available as Hampton Research Crystal Screen 1 and 2, shown in Tables 2.1 and 2.2 (Jancarik *et al.* 1991; Cudney *et al.* 1994). Promising results are then optimized to improve crystal formation.

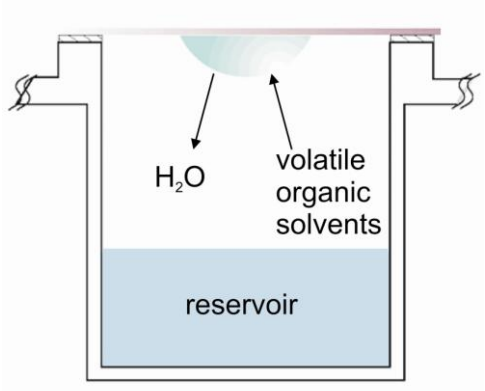


Figure 3.4: Hanging drop vapor diffusion method.
Adapted from: www.doe-mbi.ucla.edu/.

The vapor diffusion technique is often applied in crystallization trials. In this study, crystals were grown in 24-well culture plates using the hanging drop method, shown in Figure 3.4. On a siliconized glass plate 1 to 2 µl each of protein and reservoir solution were mixed together in varying ratios of 1:2, 1:1, and 2:1. The plate was then inverted and placed on

top of a well, containing 700 µl of reservoir solution. The covered well was sealed using vacuum grease preventing the setup from drying out. In the following, the protein mixture in the drop was slowly saturated by losing water molecules which evaporate into the higher concentrated reservoir solution, or by taking up volatile organic solvents from the reservoir until the equilibrium state is reached. This setup is very sensitive to temperature changes; therefore the plates were stored at a constant temperature of 4 or 20 °C.

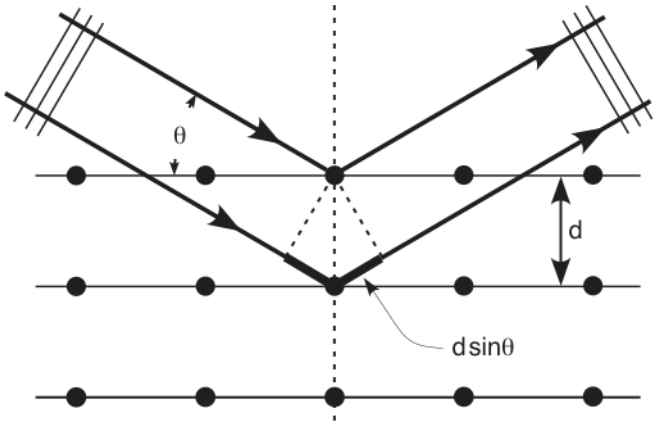
3.3.2 Diffraction data collection

Diffraction data, with a sufficient resolution for determination of the atom positions and arrangement within a protein crystal, depend entirely on the wavelength of the used electromagnetic radiation. X-rays range in wavelength from 0.1 to 10 Å. Generally a wavelength of approximately 1 Å is applied for diffraction experiments which is compare-able to the spacing between the scatterers in a crystal. Electromagnetic radiation with a continuous spectrum of X-ray energies can be obtained from electron storage rings such as BESSY, Berlin (Heinemann *et al.* 2003). The wavelength of radiation can be fine-tuned by using monochromators which is valuable for multi-wavelength anomalous dispersion (MAD) phasing.

The high intensity of synchrotron radiation requires data collection at a temperature of approximately 100 K to reduce radiation damage of the crystal. This allows collection of a complete set of X-ray diffraction data from a single crystal of high quality and resolution. Also, the signal-to-noise ratio is enhanced by reducing the atomic motion at such low temperatures. However, protein crystals have a high solvent content and, therefore, can include a high fraction of water. Flash freezing of an untreated crystal in liquid nitrogen often results in ice-crystal formation, which can instantly destroy the protein crystal or give rise to ice rings on the detected diffraction pattern, overlapping with many important data spots. Prior to crystal-freezing, the mother liquor is therefore tested and, if necessary, optimized by addition of cryo-protectants such as glycerol. The crystal is then dipped into a suitable cryo-solution, mounted in a nylon loop attached to a copper pin and flash-frozen in liquid nitrogen.

Crystals are highly periodic. They are composed of many unit cells repeating indefinitely in all three dimensions *a*, *b* and *c*, with defined angles *α*, *β* and *γ*. The protein molecules are fixated in the same orientation producing a regular array of scatterers. The incoming X-rays are altered in their direction depending on the positions of the encountered electrons relative to each other. They can constructively add up in phase producing a strong signal, or even cancel each other out, a relationship which is given by Bragg's law in equation

3.16 (Figure 3.5). Atoms of the crystal lattice are placed on sets of evenly spaced parallel layers intersecting the unit cell which are described by the Miller indices *h*, *k* and *l* with a spacing *d*. For example, the (3 0 0) planes divide the unit cell edge *a* three times, while the *b* and *c* edges are not intersected. Constructive interference only occurs between X-rays diffracted from adjacent layers, if the path length difference is equal to their wavelength λ or an integer multiple of it. Some other rules can be derived as well. First, with an increase of wavelength, the diffraction pattern is expanded and less detail is revealed at the same scattering angle. Second, the minimal distance between lattice planes constitutes the highest possible resolution, with a maximal value of $d_{max} = \lambda/2$. Most important, the law defines the reciprocal relationship between distance and diffraction angle in that the diffraction pattern is more sensitive to objects that are closer together in the crystal.



$$2d \sin\theta = n\lambda \qquad (3.16)$$

d = distance of diffracting planes
θ = diffraction angle
n = any integer
λ = wavelength

Figure 3.5: Bragg's law. Schematic representation of the diffraction of incoming X-rays on the lattice planes. Representation was taken from: http://upload.wikimedia.org/.

The reciprocal relationship can be visualized spanning a Ewald sphere (radius = 1/λ) around the crystal (Figure 3.6). For an orthorhombic, tetragonal or cubic cell the reciprocal space can be constructed with its three edges *a**, *b** and *c** (1/*a*, 1/*b* and 1/*c*) and its angles *α**, *β** and *γ** (180°-*α*, 180°-*β* and 180°-*γ*). In the center (C) at the origin of the crystal the incoming X-ray is diffracted, and leaves the sphere at the origin of the reciprocal space (O) as well as each lattice point (P) that is placed on the surface of the sphere. The diffraction vector *d** is the reciprocal distance vector with a length of 1/*d*. Hence, the two-dimensional diffraction pattern is a projection of the reciprocal lattice points (P) on the surface of the Ewald sphere, where each lattice point corresponds to one lattice plane described by the indices *h k l*. An increasing value of *h k l* is coupled with a decrease in distance between the lattice planes and a higher diffraction angle, resulting in reflections at higher resolution.

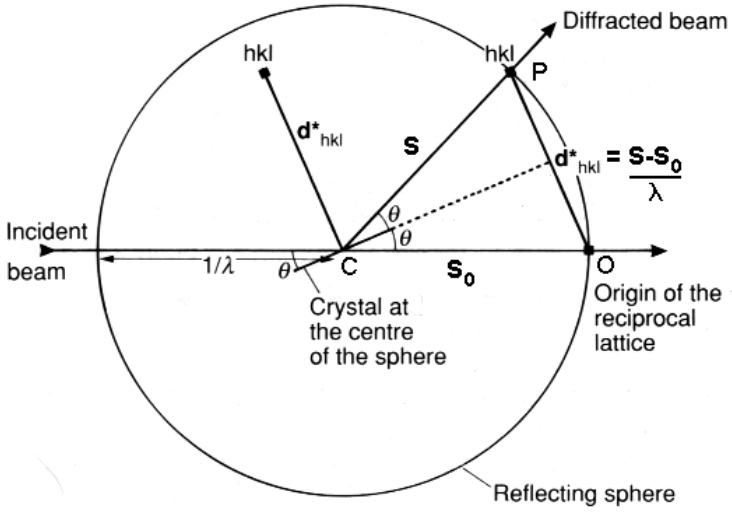


Figure 3.6: The Ewald sphere. Schematic representation of the reciprocal relationship between a crystal and its diffraction pattern. Representation was taken from: http://capsicum.me.utexas.edu/.

During the diffraction experiment, the crystal is mounted on a goniometer, positioning it precisely within the X-ray beam. X-rays are scattered into a pattern of spots or reflections, and the intensities of these reflections are detected with a fast readout detector such as a charge-coupled device (CCD) detector. As the crystal is gradually rotated, moving the reciprocal lattice through the Ewald sphere, new reflections appear while others disappear. The peaks at small angles correspond to low-resolution data, whereas those at high angles represent high-resolution data. Many images are required to cover the full reciprocal space, with fewer images needed for crystals with higher symmetry.

3.3.3 Data processing

To convert the diffraction patterns into a three-dimensional model all spots are indexed by assigning the indices *h k l* to all individual reflections. After indexing, the unit cell parameters can be obtained giving the space group of the crystal. The data are then scaled and merged. Meaning, reflection intensities that were detected several times, or follow Friedel's law where the intensities of the *h k l* and the *-h -k -l* lattice planes are identical, or which are related *via* crystallographic symmetry, are averaged. This enhances the correctness of the diffraction data. The data quality is given by the R_{sym} value (equation 3.17).

$$R_{sym} = \frac{\sum_{hkl} \sum_{i} \left| \overline{I(hkl)} - I_i(hkl) \right|}{\sum_{hkl} \sum_{i} I_i(hkl)} \qquad (3.17)$$

$I_i(hkl) =$ Intensity of single reflection

$\overline{I(hkl)} =$ Intensity of averaged reflection

It was proven that data redundancy influences the R_{sym} value effectively, therefore a more accurate and reliable value, the R_{meas} (equation 3.18) was introduced (Diederichs and Karplus 1997). A single reflection is now independent of averaging.

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n_h}{(n_h - 1)}} \sum_i \left| \overline{I(hkl)} - I_i(hkl) \right|}{\sum_{hkl} \sum_i I_i(hkl)} \tag{3.18}$$

All data sets were indexed, integrated and scaled using the *XDS* program package (Kabsch 1993). The input file XDS.INP was prepared and after indexing following the *XYCORR*, *INIT*, *COLSPOT* and *IDXREF* jobs, the obtained output file included the unit cell parameters and the correct space group. The reflections were then integrated and scaled by running *DEFPIX*, *XPLAN*, *INTEGRATE* and *CORRECT*. *XDSSCALE* performs the final scaling and merging steps, and *XDSCONV* converts the output file for the use within the *CCP4* suite (Collaborative Computational Project 1994).

3.3.4 The electron density

The two-dimensional diffraction images can be converted into a three-dimensional model of the electron density using Fourier transformation. The structure factor $\mathbf{F}(hkl)$ describes each reflection of the diffraction pattern as a Fourier transform of the atomic scattering factors f_j of each atom *j* within the crystallographic unit cell.

$$\mathbf{F}(hkl) = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)} \tag{3.19}$$

$\mathbf{F}(hkl)$ = structure factor of a diffracted X-ray
f_j = atomic scattering factor of atom j
h, k, l = reflection indices in reciprocal space
x_j, y_j, z_j = coordinates in real space of atom j

The structure factor defines the integral of the electron density over a specific volume element.

$$\mathbf{F}(hkl) = \int_V \rho(xyz) e^{2\pi i(hx+ky+lz)} dV \tag{3.20}$$

$\rho(xyz)$ = electron density
V = cell volume

Hence, the electron density can be obtained with the inverse Fourier transform of the structure factors.

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l \mathbf{F}(hkl) e^{-2\pi i(hx+ky+lz)} = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| e^{i\alpha(hkl)} e^{-2\pi i(hx+ky+lz)} \tag{3.21}$$

$|F(hkl)|$ = amplitude
α = phase shift

The structure factor or Fourier transform is a complex number consisting of an amplitude $|F(hkl)|$ and a phase α (3.21). The square of the structure factor amplitude is proportional to the intensity of each reflection (3.22). To obtain an interpretable electron density map, the amplitude and the phase must be known. However, the phase cannot be directly determined during a diffraction experiment.

$$\sqrt{|I(hkl)|} \sim |F(hkl)| \tag{3.22}$$

3.3.5 Obtaining phases

After assigning the detected individual reflection intensities the process of structure solving can proceed in many different ways. Various methods can be applied to overcome the phase problem:

- Isomorphous replacement (SIR, MIR)
- Anomalous dispersion (SAD, MAD)
- Molecular replacement (MR)
- Direct methods

In single or multiple isomorphous replacement (SIR, MIR) phases are estimated from the differences between the amplitudes of the diffraction maxima from the crystals of the native macromolecule and its heavy-atom derivatives. The direct phase determination using anomalous diffraction information involving the measurement of diffraction data at several different wavelengths is called multiple-wavelength anomalous diffraction (MAD) phasing. Isomorphous replacement and anomalous scattering can also be combined (SIRAS, MIRAS). MAD phasing offers several significant advantages compared to MIR and other phasing techniques. First, all data can be determined using the same crystal. Therefore data quality is improved by removing many sources of systematic errors. Second, anomalous scattering remains strong at high resolution, at which the resulting phase angles are more accurate. Third, a single experiment leads to an interpretable electron density map of generally higher quality. Hence, the initial model can be determined much faster and is more accurate.

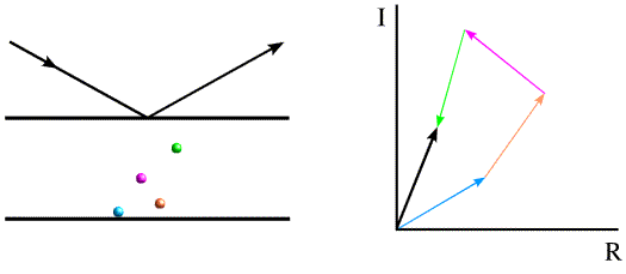
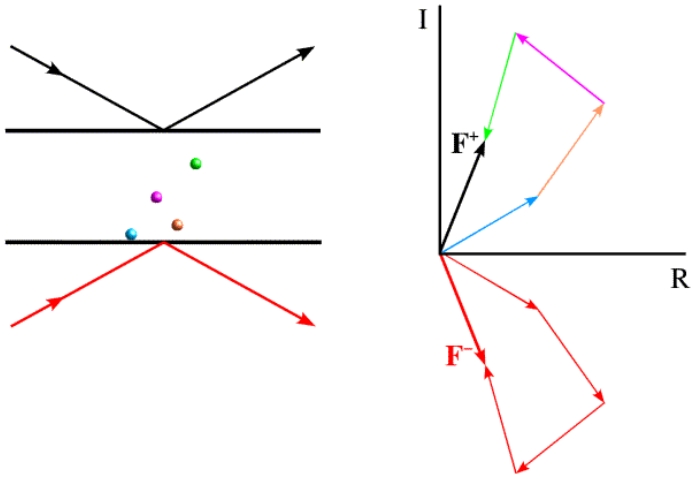


Figure 3.7: The structure factor. The structure factor describes the overall diffracted wave from a set of lattice planes. The different contributions of the four light atoms are visualized by matching colors. The vector of the structure factor (black) is the overall sum of the colored vectors (http://www-structmed.cimr.cam.ac.uk/).

In Figures 3.7 and 3.8 two lattice planes with four light atoms in between are visualized. The relative phase shift ($0 \leq \alpha \leq 2\pi$) is coupled to the relative distance of the four atoms from the planes that define a phase angle of zero. The vector of the overall diffracted X-ray is the sum of the vectors of the four atoms. Most scatterers within the crystal will diffract the X-rays with a relative phase of zero. Friedel's law (equation 3.23) describes that sets of reflections indicate a very similar diffraction event with the same angles of incident and diffracted waves. In Figure 3.8, the diffracted X-rays from one set of lattice planes (black arrows) will give rise to one reflection indexed using the indices *h k l* from the lattice planes, while the diffracted beam from the other set of lattice planes (red arrows) define the *-h -k -l* planes. In both cases, the four atoms will contribute with the same phase shift therefore the intensity of *h k l* is the same as that of *-h -k -l*. The diffraction image is centro-symmetric and the data can be cut in half using scaling and merging procedures.



$$|F(hkl)| = |F(\overline{hkl})| \qquad (3.23)$$

→ $F^+ = F^-$

Figure 3.8: The Friedel pairs. Representation taken from: http://www-structmed.cimr.cam.ac.uk/.

A shift in phase and amplitude is called anomalous scattering. Heavy atoms, featuring the property of anomalous scattering, can be naturally present in the molecule of interest or they can be incorporated before crystallization. The incorporation of selenomethionine into recombinant proteins is often applied (Hendrickson *et al.* 1990). For the structure determination of nucleic acids, the chemical incorporation of brominated nucleosides is equally important. 5-bromodeoxyuracil is often used as the isosteric substitute for thymine (Correll *et al.* 1997).

Anomalous scattering methods utilize the resonant behavior of heavy atoms. Is the wavelength of the incident X-rays in the vicinity of their absorption edges, the inner electrons are elevated to higher states, and when they fall back to the lower energy state, the electrons emit radiation (Hendrickson 1991). Absorption edges of carbon, nitrogen, oxygen and sulfur are not near the accessible energy range at synchrotrons. In general,

elements with atomic numbers above 20 are used. Resonance effects result in variable phase shifts of the scattered wave. They are expressed by the real and imaginary anomalous scattering factors f' and f'', which are functions of the wavelength and show significant variations near the absorption edge (equation 3.24).

$$f_{ano}(\theta,\lambda) = f^{0}(\theta) + f'(\lambda) + if''(\lambda) \tag{3.24}$$

f_{ano} = total anomalous scattering factor, depending on θ and λ
θ = angle
λ = wavelength
f^{0} = normal atomic scattering without phase shift
f' = real (dispersive) correction term
f'' = imaginary (anomalous) correction term (delay of emission leads to a phase shift of 90° against f^{0} *and* f')

The additional imaginary component of the total anomalous scattering factor is always positive leading to the breakdown of Friedel's law (Figure 3.9 and equation 3.25). However, this is not the case for centro-symmetric space groups, where Friedel's law is always true, and anomalous phasing cannot be applied for those cases.

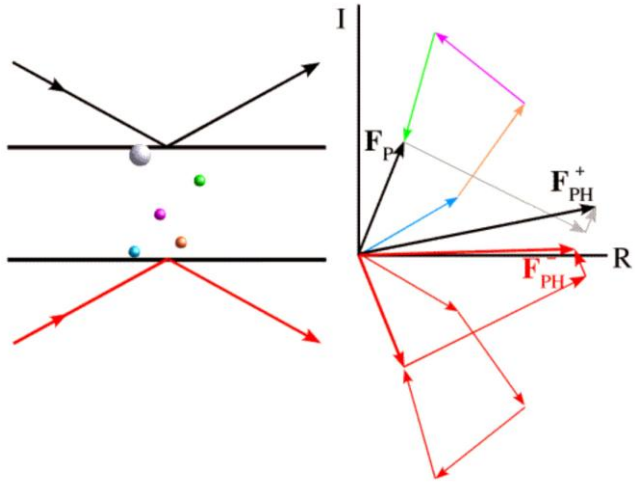


Figure 3.9: Anomalous scattering. The heavy atom (gray) with an anomalous scattering component leads to a break-down of Friedel's law. Representation taken from: http://www-structmed.cimr.cam.ac.uk/.

$$\left|F(hkl)\right| \neq \left|F(\overline{hkl})\right| \tag{3.25}$$

The Harker construction in Figure 3.10 visualizes that phase determination in multiple isomorphous replacement, using data from a native and at least two derivative crystals, is unequivocal. For MAD phasing this concept can be applied as well when using the anomalous signal as derivative. Several data sets are collected at different wavelengths, where the anomalous scattering factors f' and f'' reach extreme values. The peak data set is collected at the wavelength where f'' reaches its maximum. Since the power of

phasing is directly proportional to the value of f'', it is measured first, because radiation damage of the crystal during data collection can always limit the number of complete data sets. For the inflection point data set the wavelength where f' reaches its minimum is applied, and for the remote data set a wavelength further away from the absorption edge on the high-energy and low-energy sides is taken. To obtain these wavelengths for a protein-DNA crystal incorporating 5-bromodeoxyuracil, a polarized X-ray absorption spectrum is recorded at energy values around the theoretical Br *K* edge of the free bromine atom (E = 13474 eV, λ = 0.9292 Á), in fluorescence excitation mode. The fluorescence spectrum is transformed into f' and f'' spectra using the program *CHOOCH* (Evans and Pettifer 2001). This program uses the measured anomalous f'' term to calculate the dispersive f' term with the Kramer-Kronig transformation (equation 3.26).

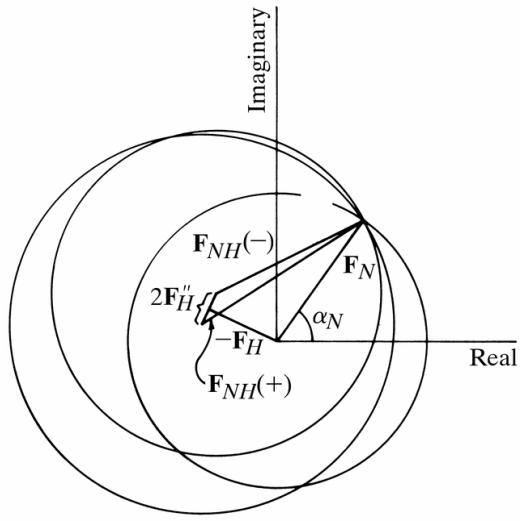


Figure 3.10: The Harker construction. Phase determination is unequivocal for MAD phasing. Following the principle of multiple isomorphous replacement using one native *N* and two derivative $NH^{(+)}$ and $NH^{(-)}$ crystals, only one phase α_N is consistent with all observations. The point where the three circles center gives the only solution for the structure factor $\mathbf{F}_N$. Representation taken from: http://it.iucr.org/.

$$f'(E_0) = \frac{2}{\pi}\int_0^{\infty} \frac{Ef''(E)}{E_0^2 - E^2}\, dE \qquad (3.26)$$

The expected structure factor ${}^{\lambda}\mathbf{F}(\mathbf{h})$ at wavelength λ for a reflection $\mathbf{h}(hkl)$ is defined as:

$${}^{\lambda}\mathbf{F}(\mathbf{h}) = {}^{0}\mathbf{F}_P(\mathbf{h}) + {}^{\lambda}\mathbf{F}_H(\mathbf{h}) \quad \rightarrow \quad {}^{\lambda}\mathbf{F}(\mathbf{h}) = {}^{0}\mathbf{F}_{PH}(\mathbf{h}) + {}^{\lambda}\mathbf{F}'_H(\mathbf{h}) + i\,{}^{\lambda}\mathbf{F}''_H(\mathbf{h}) \qquad (3.27)$$

$\mathbf{F}(\mathbf{h})$ = structure factor
P = normal scatterer
H = anomalous scatterer
PH = total number of scatterers

If all anomalous scatterers are of one kind, the following can be assumed:

$${}^{\lambda}\mathbf{F}'_{\mathrm{H}} = \frac{f'(\lambda)_0}{f^0}\mathbf{F}_{\mathrm{H}} \quad \text{and} \quad {}^{\lambda}\mathbf{F}''_{\mathrm{H}} = \frac{f''(\lambda)_0}{f^0}\mathbf{F}_{\mathrm{H}} \qquad (3.28)$$

The structure factors of Friedel pairs $|^{\lambda}F_{obs}(+h)|$ and $|^{\lambda}F_{obs}(-h)|$ can be expressed in wavelength-dependent and -independent terms (equation 3.29). Using the values for f' and f'', the phases α can be calculated. For each reflection two possible equations are obtained. The accuracy of the solution depends on the value of the differences between the anomalous scattering factors which are most prominent in the peak data set.

$$\left|{}^{\lambda}F_{obs}(\pm h)\right|^2 = \left|{}^{0}F_{HP}\right|^2 + a(\lambda)\left|{}^{0}F_{H}\right|^2 + b(\lambda)\left|{}^{0}F_{HP}\right|\left|{}^{0}F_{H}\right|\cos({}^{0}\alpha_{HP} - {}^{0}\alpha_{H}) \pm c(\lambda)\left|{}^{0}F_{HP}\right|\left|{}^{0}F_{H}\right|\sin({}^{0}\alpha_{HP} - {}^{0}\alpha_{H}) \tag{3.29}$$

$$\text{with:} \quad a(\lambda) = \frac{f'^2 + f''^2}{f^{0^2}} \qquad b(\lambda) = 2\frac{f'}{f^0} \qquad c(\lambda) = 2\frac{f''}{f^0}$$

The inverse Fourier transform of the structure factors can be used to obtain the electron density of the macromolecule (equation 3.20). If only experimentally measured reflection intensities are inserted, this is called a Patterson function (equation 3.30). This method results in a map whose maxima correspond to vectors between atoms. If the structure of interest is not as complicated and large as a macromolecule, it can be deconvoluted into atomic positions. Hence, the Patterson method can be applied to obtain interpretable maps of the anomalous scatterers' substructure. Differences between Friedel pairs are used to calculate a difference Patterson map.

$$P(u,v,w) = \frac{1}{V}\sum_{hkl}\left|F_{hkl}\right|^2 e^{-2\pi(hu+kv+lw)} \tag{3.30}$$

The *HKL2MAP* program suite (Pape and Schneider 2004) was used to derive phases from the experimental data. The $\mathbf{F}_H$ values were calculated with *SHELXC* (Sheldrick 2008), and the anomalous scattering substructure was determined with *SHELXD* (Schneider and Sheldrick 2002). Finally, the phases were calculated with *SHELXE* (Schneider and Sheldrick 2002), and the electron density map was determined. The electron density map was improved using the program *DM* (Cowtan 1994), and eventually a first protein model could be built by *warpNtrace* (Lamzin *et al.* 2001).

3.3.6 Model building and refinement

The first protein model obtained in this way was manually extended, reoriented and improved. Automated model building was only accomplished for protein structures; therefore, a DNA model was inserted manually into the electron density. These processes were guided by using difference electron density maps. The $2F_o$-F_c map was contoured at +1σ. It was derived from the experimentally measured data.

$2F_o$-F_c map: $$\rho(xyz) = \frac{1}{V}\sum_h \sum_k \sum_l \left\| 2F_{obs} \right| - \left| F_{calc} \right\| e^{-2\pi i(hx+ky+lz)+i\alpha_{calc}} \quad (3.31)$$

The F_o-F_c maps were contoured at +3σ for positive or -3σ for negative difference density. Both maps display differences between the experimentally measured X-ray data and the structure amplitudes derived from the crystallographic model.

F_o-F_c map: $$\rho(xyz) = \frac{1}{V}\sum_h \sum_k \sum_l \left\| F_{obs} \right| - \left| F_{calc} \right\| e^{-2\pi i(hx+ky+lz)+i\alpha_{calc}} \quad (3.32)$$

Model adjustment in real space is followed by refinement that optimizes atomic positions in reciprocal space. The structure factor amplitudes of the refined model are compared to the experimental data, giving the *R* factor (equation 3.33). The *R* values monitor the improvement of the refinement. R_{work} compares the current refined model with all X-ray data. R_{free} is calculated on a small subset of the data (5 – 10%) that is not used in the refinement process. The independence of the R_{free} is used to prevent overfitting of the data.

$$R_{work/\,free} = \frac{\sum_{hkl} \left\| F_{obs} \right| - \left| F_{calc} \right\|}{\sum_{hkl} \left| F_{obs} \right|} \quad (3.33)$$

During refinement the model was restrained using force-field methods based on defined parameters for covalent bond lengths, bond angles, torsion angles, and planar groups which were weighted differently using n_1 – n_5 (Kini and Evans 1991). The energy function given in equation 3.34 represents energy values of a stereochemical (E_{stereo}) and an X-ray (E_{xray}) term. The total energy value has to decrease during a refinement run.

$$E_{total} = E_{stereo} + E_{xray} = n_1\left(n_2 E_{bond\,angles} + n_3 E_{bond\,dist} + n_4 E_{torsion\,angles} + n_5 E_{planar\,groups}\right) + E_{xray} \quad (3.34)$$

Refinement of atom displacement is monitored by the Debye-Waller term. This term, the *B* factor, indicates the extent of oscillation or vibration of each individual atom around its specified position within the model. This step can be carried out isotropically or anisotropically, describing the displacement of atoms using spherical or ellipsoid models, respectively.

$$B_j = 8\pi^2 U_j^2 \quad (3.35)$$

B_j = Debye-Waller factor of atom *j*
U_j^2 = mean square displacement of atom *j*

An additional weighting term for each atom *j* is introduced into the structure factor equation 3.19, which corrects the atomic scattering factor.

$$\mathrm{F}(hkl) = \sum_j f_j e^{-\frac{B_j \sin(\theta)^2}{\lambda^2}} e^{2\pi i(hx_j + ky_j + lz_j)} \tag{3.36}$$

Restrained anisotropic refinement describes atomic motion better than an isotropic refinement. However, it is only appropriate for structures with severe conformational disorder or for crystals that diffract to better than 1.5 Å (Gros *et al.* 1990; Clarage and Phillips 1994). For refining the atomic displacement anisotropically, additional parameters are required to define a thermal ellipsoid. Translational, librational and screw-motion parameters describe the displacement. To limit the number of refinement parameters they can be grouped together in TLS groups (Winn *et al.* 2001).

The model was visualized and manually adjusted with the graphics program *O* (Jones *et al.* 1991), and the electron density maps were obtained from the program *FFT* (Collaborative Computational Project 1994). Refinement with the program *REFMAC5* (Murshudov *et al.* 1997) further improved the model. *warpNtrace* (Lamzin *et al.* 2001) was used for automated model building, and water molecules were added to the model using *ARP/wARP* (Lamzin *et al.* 2001). TLS groups were assigned for the last refinement steps using the *TLSANAL* program (Collaborative Computational Project 1994). The final structure model was validated with programs such as *PROCHECK* (Laskowski *et al.* 1993) and *WHATCHECK* (Hooft *et al.* 1996).

4 Results

4.1 Structure determination

4.1.1 Recombinant expression of *korA*

The coding region for the KorA protein (residues 1–101) was cloned into the bacterial expression vector pMS470Δ8 as described by Balzer *et al.* (1992). The *korA* expression plasmid pJM101a-3 was kindly provided by Dr. Erich Lanka (Max-Planck Institute for Molecular Genetics, Berlin). The mutant variants R48A, Q53A, R48A/Q53A, and Q53E were generated using the QuikChange mutagenesis protocol (3.1.5). *E. coli* BL21(DE3) carrying the appropriate wild-type or mutant *korA* expression plasmid were cultured in LB medium containing 100 µg ml^{-1} ampicillin at 37 °C. Transcription in cells was induced at OD_{600} = 0.8 by addition of 1 mM IPTG, and growth was continued for 4 h.

4.1.2 Purification of KorA

In the following text, the purification of wt KorA and its mutant variants is described. If not stated otherwise, all variants were produced accordingly. After harvesting, cells were resuspended in lysis buffer (2.4) and disrupted using a French Press twice at 1100 psi. In the first purification steps, the lysate was cleared from nucleic acids according to the purification protocol for PEI precipitation (3.2.3), and fractionated ammonium sulfate precipitation (ASP, see 3.2.4) partially removed unwanted proteins. The 30 – 50% fraction of ASP, enriched in KorA protein, was dialyzed in dialyzing buffer (2.4) for 6 h. All steps were analyzed by SDS PAGE, as shown in Figure 4.1.

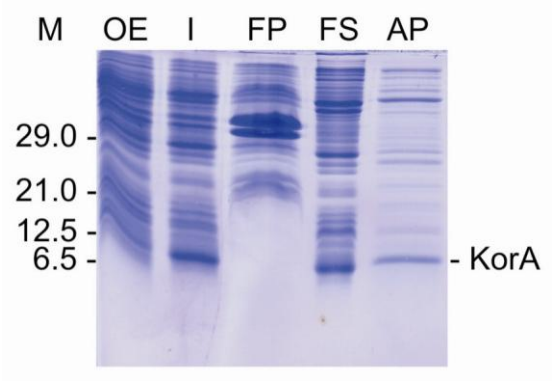


Figure 4.1: SDS PAGE showing the first purification steps of KorA. M: marker lanes (molecular weight of marker proteins is given in kDa); OE: overnight expression; I: IPTG induced; FP: French Press insoluble fraction; FS: French Press soluble fraction; AP: 30-50% fraction of ammonium sulfate precipitation.

KorA was further purified using affinity chromatography (3.2.6). After binding, the affinity matrix was washed extensively with wash buffer (2.4), before a linear salt gradient of 0 - 1 M NaCl over a volume of 200 ml resulted in elution of all bound proteins, as seen in Figure 4.2. wt KorA, Q53A, and Q53E elute between 0.5 – 0.6 M of NaCl, while the mutants R48A and R48A/Q53A elute earlier between 0.3 – 0.4 M NaCl.

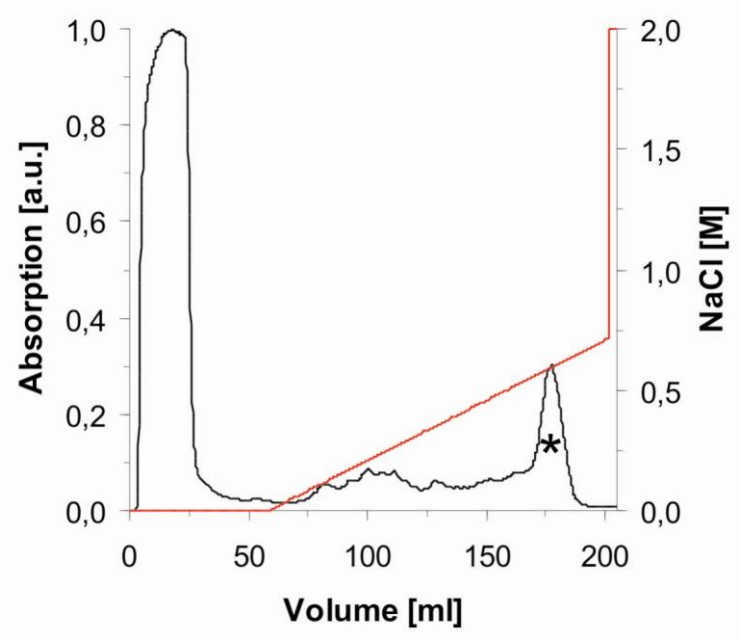


Figure 4.2: Affinity chromatography of partially pure KorA. The HiTrap Heparin column was loaded with a protein sample containing wt KorA. After washing, the bound proteins were eluted by a linear salt gradient (red line). KorA protein (marked with an asterisk) elutes at 0.5 - 0.6 M of NaCl. The protein concentration was monitored measuring the absorbance at 280 nm.

KorA protein, with a pI value of 9.6, was then specifically bound to a cation-exchange column (3.2.7). A linear salt gradient of 0 – 1.0 M of NaCl over a volume of 50 ml resulted in elution of KorA at 0.2 – 0.4 M of NaCl, as seen in Figure 4.3.

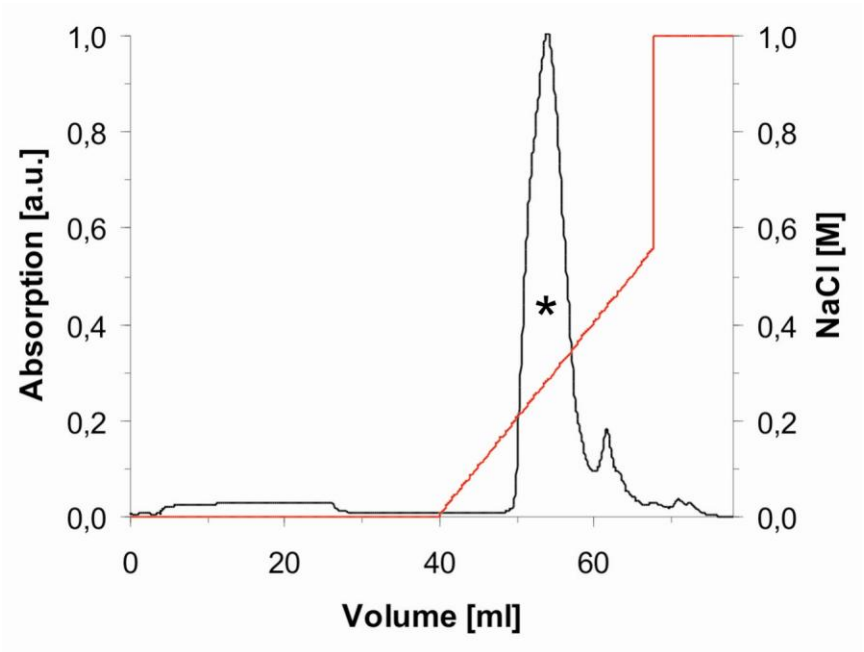


Figure 4.3: Cation-exchange chromatography of KorA. The Mono S column was loaded with a protein sample containing KorA. After washing, the bound proteins were eluted by a linear salt gradient (red line). KorA protein (marked with an asterisk) elutes at 0.2 - 0.4 M of NaCl. Protein concentration was monitored measuring the absorbance at 280 nm.

In the final purification step the proteins were subjected to size-exclusion chromatography (3.2.8). This step was required to remove some high-molecular-weight impurities still present in the sample. The gel filtration column was equilibrated with gel filtration buffer (2.4) prior to loading the protein sample. KorA migrates as a single peak and elutes at a volume corresponding to a homodimer. Protein containing fractions were concentrated in a 5K-cutoff Ultrafree Millipore membrane.

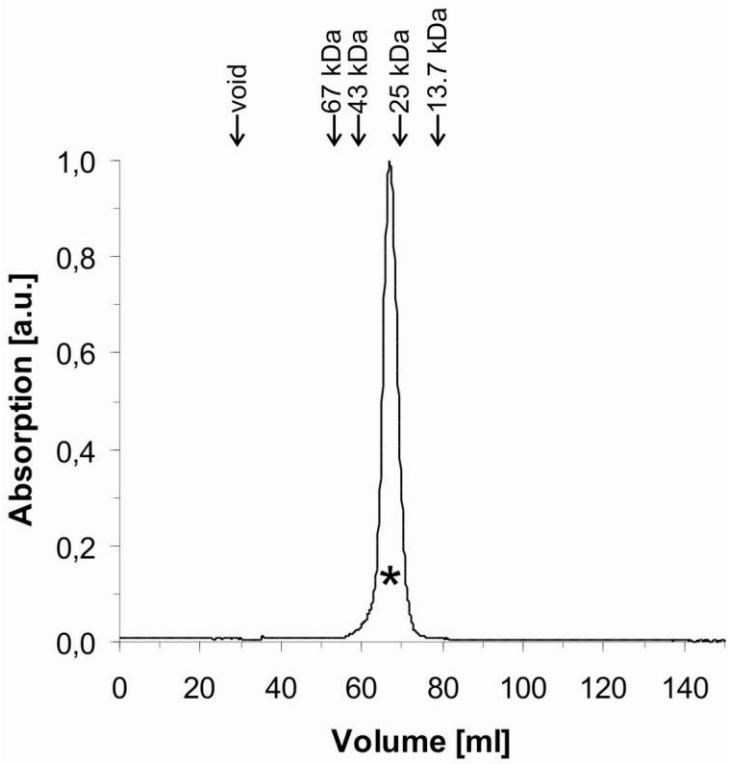


Figure 4.4: Gel filtration elution profile of KorA. KorA protein (marked with an asterisk), with a molecular mass of 11.3 kDa, elutes as a stable dimer from a calibrated Superdex 75 (16/60) column. Protein concentration was monitored measuring the absorbance at 280 nm. The molecular weight standards of the column calibration are indicated by black arrows on the elution profile.

All eluates were collected and analyzed by SDS PAGE (4.5). The purified full-length wt KorA protein was stable and dimeric in solution as confirmed by analytical ultracentrifugation (Figure 4.6).

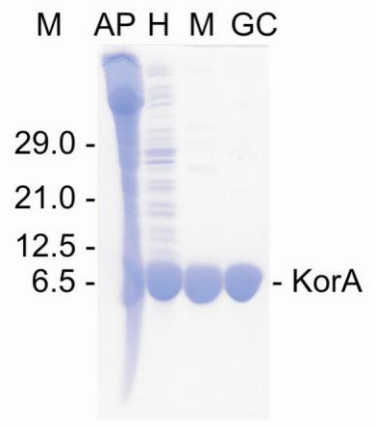


Figure 4.5: SDS PAGE showing the final purification steps of KorA. M: marker lanes (molecular weight of marker proteins is given in kDa); AP: ammonium sulfate precipitation (30 – 50% fraction); H: after affinity chromatography with a Heparin column; M: after cation-exchange chromatography with a Mono S column; GC: after gel filtration with a Superdex 75 column.

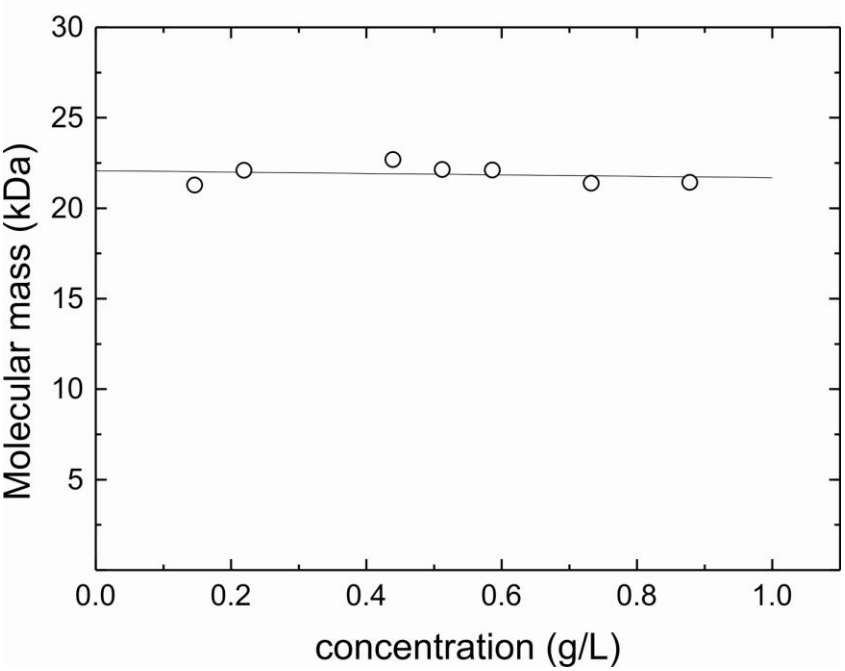


Figure 4.6: Mass determination of KorA in solution using analytical ultracentrifugation. The molecular mass extrapolated to infinite dilution is approximately 22.1 kDa, proving the stable dimeric state of KorA in solution. The molecular weight of the KorA monomer is 11.3 kDa.

4.1.3 Formation of the KorA-O_A* complex

The choice of the 18-bp oligonucleotide O_A for crystallization was made during my diploma thesis, when I started working on this project. Purified KorA protein was incubated with synthetic oligonucleotides of various lengths (12-bp, 14-bp, 16-bp and 18-bp), and analyzed on native PAGE. The tightest binding was observed for the two longer DNA constructs. The sequence was taken from the O_A1 site at the *korA* promoter, which was shown to bind to KorA with the highest affinity (Table 1.2). Since no structural homolog of KorA was known, bromine atoms were incorporated in the DNA molecule, allowing MAD phasing experiments.

The HPLC-purified 18-bp oligonucleotide O_A* (2.5.3), was used for complex formation with KorA. Outside of the 12-bp operator sequence, three thymine residues were replaced by 5-bromodeoxyuracil within one DNA strand so as to avoid any interference in KorA binding to the consensus sequence by the added bromine atoms. The light-sensitive brominated DNA was kept in the dark at all time. Strands were mixed in 1:1 ratio, annealed and purified on a gel filtration column (Figure 4.7), equilibrated with gel filtration buffer (2.4). KorA and DNA were mixed in a 2.5:1 molar ratio, incubated at 37 °C for 20 min, and subjected to size-exclusion chromatography to separate the 2:1 KorA-O_A* complex of 33.6 kDa from free DNA and protein. Afterwards, the complex was concentrated to 3.8 mg/ml.

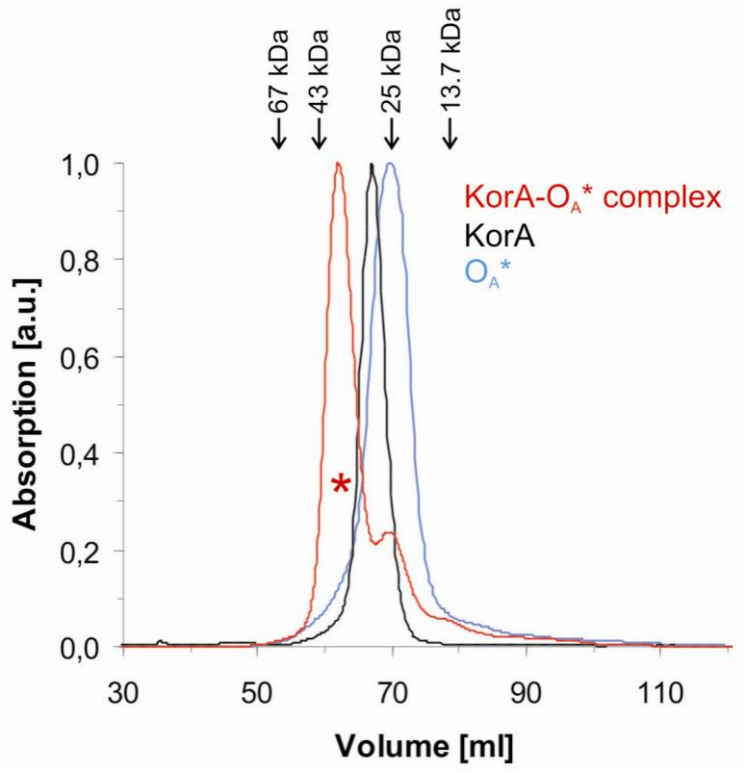


Figure 4.7: Gel filtration elution profile of the KorA-O_A* complex. The 2:1 KorA-O_A* complex of 33.6 kDa (marked with a red asterisk) elutes as a stable complex from a calibrated Superdex 75 (16/60) column (red). The elution profiles of the 18-bp O_A* oligonucleotide (blue) and dimeric KorA protein (black) are shown for comparison. The protein concentration was monitored measuring the absorbance at 280 nm. The molecular weight standards of the column calibration are indicated by black arrows on top of the elution profile.

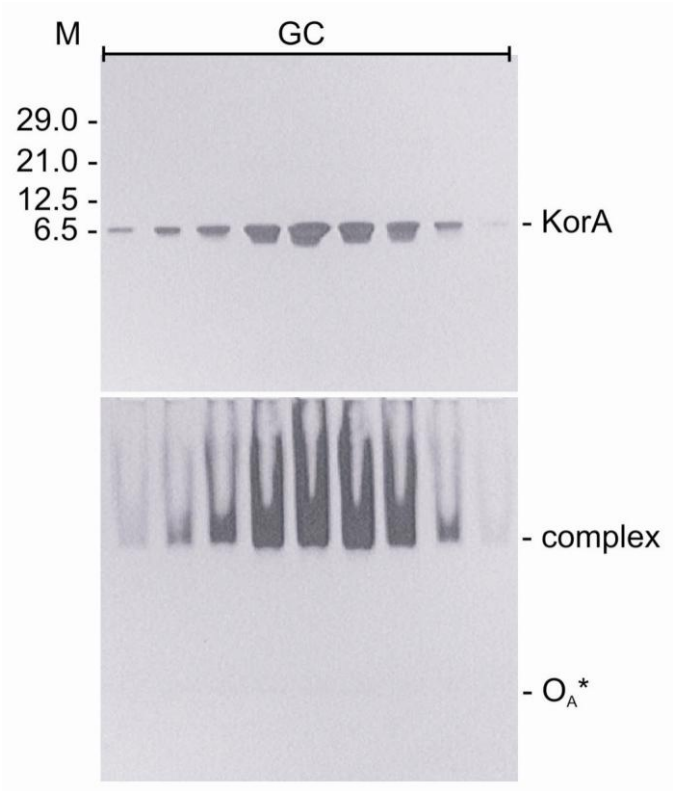


Figure 4.8: PAGE evaluation of the gel filtration run of KorA-O_A* complex. Top, SDS PAGE: M: marker lanes (molecular weight of marker proteins is given in kDa); GC: pooled fractions of the complex showing the presence of KorA. Bottom, native PAGE: GC: pooled fractions corresponding to the gel above; complex: KorA-O_A* complex; O_A*: unbound 18-bp oligonucleotide. (All gels were silver stained.)

4.1.4 Crystallization of the KorA-O_A* complex

The full-length transcriptional repressor KorA bound to the pseudo-symmetric 18-bp oligonucleotide, O_A*, was crystallized. Crystallization conditions were identified by the hanging-drop vapor diffusion method. Initial screens with Hampton Crystal Screen 1 and 2 (see 2.4) already produced crystals in many conditions. The best crystals of the KorA-O_A* complex were obtained by mixing 2 µl of KorA-O_A* complex (3.8 mg/ml) with 1 µl of reservoir solution containing 0.5 M $(NH_4)_2SO_4$, 0.1 M Na^+ citrate (pH 5.6), and 1.0 M Li_2SO_4. Monoclinic crystals grew in 3 – 5 days at 20 °C in the dark (Figure 4.9). KorA protein without DNA did not crystallize under any condition tested.

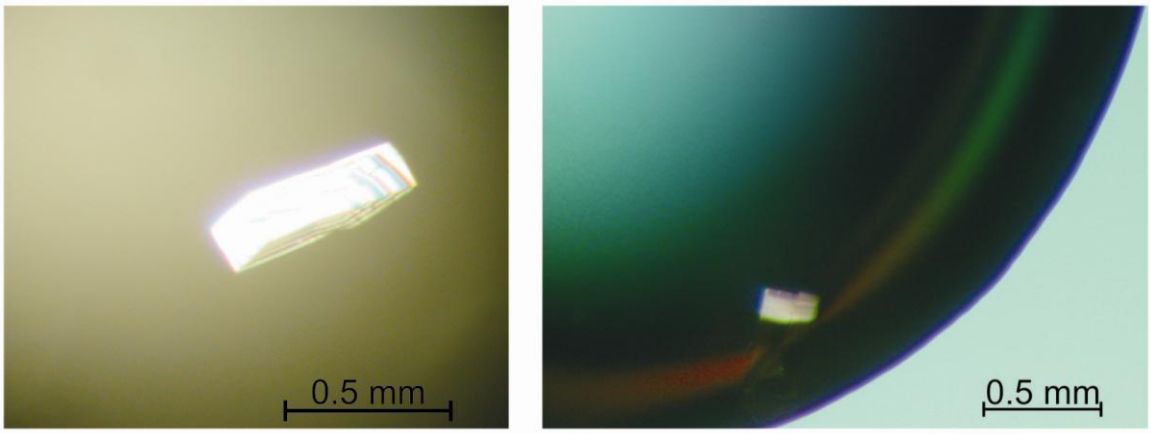


Figure 4.9: Crystals of the KorA-O_A* complex. Crystals grown in 0.5 M $(NH_4)_2SO_4$, 0.1 M Na citrate pH 5.6, and 1.0 M Li_2SO_4. The crystal shown on the right was used for data collection.

4.1.5 Data collection

The crystal was frozen in liquid nitrogen. The solution for cryo-protection contained reservoir solution with 15% (v/v) of glycerol. A two-wavelength MAD experiment was carried out at 100 K, in the dark, using synchrotron radiation at beamline BL 14.1 (Heinemann *et al.* 2003) at BESSY (Berlin) on a fast scanning 225 mm CCD-mosaic detector from MAR Research (Norderstedt). To determine the optimal wavelengths for anomalous data collection, a fluorescence scan of the crystal near the bromine edge was performed.

The program *CHOOCH* (Evans and Pettifer 2001) was used to calculate the anomalous scattering curves. The results are shown in Figure 4.10 and Table 4.1. It was not possible to obtain a wavelength at a reasonable inflection point, and therefore data sets to 1.85 Å were collected at two more wavelengths within the range of the theoretically expected inflection point (personal communication with Dr. Uwe Müller, BESSY Berlin). The crystal was gradually rotated in angular increments of 1° covering 360°. At every step an image was recorded.

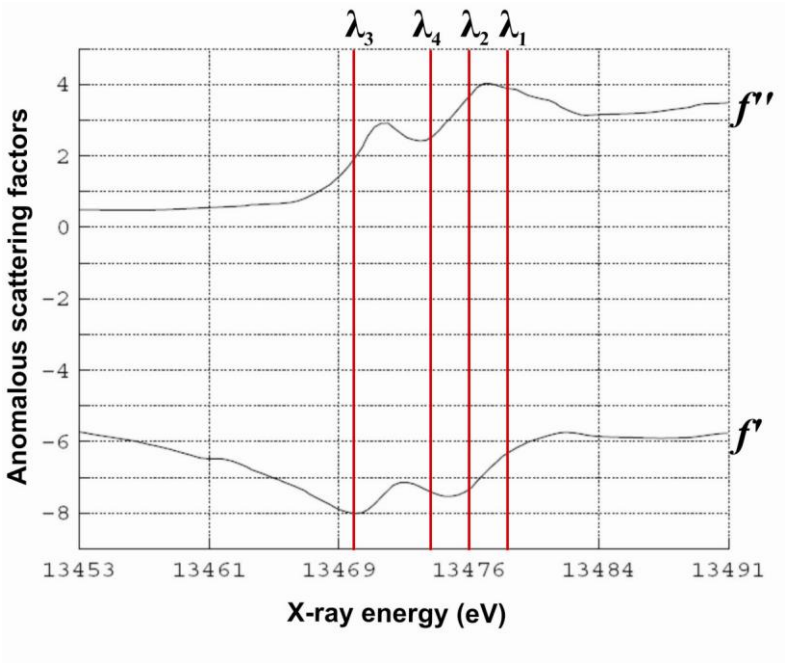


Figure 4.10: Anomalous scattering curves. The obtained f' and f'' values using the program *CHOOCH* were used to define the wavelengths λ_1 for the peak and λ_3 for the inflection point. Two more wavelengths, λ_2 and λ_4, were selected manually at possible inflection points.

	X-ray energy (eV)	wavelength (Å)	f'	f''
Peak[a]	13477.99	λ_1 = 0.91991	-6.7	4.0
Inflection 1[b]	13476.00	λ_2 = 0.92004	-7.4	3.6
Inflection 2[a]	13470.00	λ_3 = 0.92045	-8.0	2.0
Inflection 3[b]	13474.00	λ_4 = 0.92018	-7.5	2.5

[a]Obtained by the program *CHOOCH*.
[b]Additionally selected for data set collection.

Table 4.1: Anomalous scattering factors at peak and inflection wavelengths.

The X-ray diffraction data were processed and scaled using the *XDS* program suite (Kabsch 1993). The data collection statistics are listed in Table 4.2. The KorA-O_A* complex crystallized in space group $P2_1$ with unit cell parameters of a = 44.83 Å, b = 115.45 Å, c = 49.88 Å, β = 114.0°. The asymmetric unit contains one copy of the 33.6 kDa complex, with a solvent content of 65%.

Data set	**Peak**	**Inflection 1**	**Native-like merged peak**
Wavelength (Å)	0.91991	0.92004	0.91991
Resolution range (Å)[a]	50-1.96	50-1.96	50-1.96
	(2.08-1.96)	(2.08-1.96)	(2.08-1.96)
Total reflections[a,b]	229,922	216,313	230,648
	(19,830)	(18,556)	(20,085)
Unique reflections[a]	60,942	60,954	31,085
	(7,376)[b]	(7,347)[b]	(3,817)
Completeness (%)[a]	92.6 (68.4)	92.5 (68.0)	93.3 (70.1)
$<I / \sigma(I)>$[a]	32.2 (12.7)	37.1 (11.4)	42.5 (17.4)
R_{sym} (%)[a]	2.5 (6.5)	2.2 (7.5)	2.7 (7.4)
R_{meas} (%)[a,c]	2.9 (8.3)	2.5 (9.7)	2.9 (8.2)

[a]Values in parentheses are for the highest resolution shell.
[b]Friedel pairs not merged.
[c]Multiplicity-corrected R_{sym} as defined by Diederichs and Karplus (1997).

Table 4.2: Data collection statistics.

4.1.6 Structure analysis and refinement

The structure of the complex containing KorA (1-101) and the 18-bp O_A* oligomer was solved by MAD phasing at a resolution of 1.96 Å. With the program suite *HKL2MAP* (Pape and Schneider 2004), the best results were obtained using the peak and inflection 1 data sets. Six heavy atom sites were located, differing in relative occupancy and anomalous contribution. The number and position of the bromine sites was the first indication of a two-fold disorder of the O_A* duplex.

The density modification program *DM* (Cowtan 1994) was used to improve the quality of the electron density map. The non-crystallographic symmetry (NCS) of the heavy atoms in the ASU permitted subsequent phase improvement by the *DM* program including automatic solvent and NCS masking as well as solvent flattening. *warpNtrace*

implemented in the *ARP/wARP* software package (Lamzin *et al.* 2001) automatically built a first protein model (chains A 4-31, A 34-54, B 25-37 and B 47-61). This model was then manually adjusted and extended in *O* (Jones *et al.* 1991). Superposition of the A and B chains of KorA in *LSQKAB* (Kabsch 1978) was used to extend the model (chains A 3-61 and B 12-62). The structural model was refined using *REFMAC5* and loose NCS restraints (Murshudov *et al.* 1997). 5% of the reflections were set aside for cross-validation, and the R_{free} value was used to monitor model refinement strategy and progress. After five runs of *warpNtrace* a DNA model was manually inserted into the positive F_o-F_c difference density and adjusted, guided by three bromine positions. The 18-mer DNA model (strands E and F) was generated with the program *MAKE-NA* (http://structure.usc.edu/make-na/). Water molecules were added to the model (chains A 3-93 and B 5-93) using *ARP/wARP*. Strong positive density was observed at the three terminal base pairs on both ends. This was satisfied by superimposing the DNA duplex in two orientations, and fitting the second double strand into the density using the three non-occupied bromine positions. The occupancy of all DNA atoms was adjusted to 0.5, and the six bromine positions were set to 0.25. 36 TLS groups were assigned using the *TLSANAL* program (Collaborative Computational Project 1994). Final TLS refinement steps resulted in R_{free} and R_{work} of 20.1% and 17.0%, respectively. The refinement statistics are given in Table 4.2, and Figure 4.11 illustrates the crystal packing of the KorA-O_A* complex. Diffraction data and atomic coordinates were deposited in the Protein Data Bank with accession code 2w7n.

The electron density of the complex is well defined. Alternative side chain conformations were observed for Glu63 (chains A and B of KorA) as well as Glu25 and Asp64 (chain B). These three solvent-exposed residues are not involved in KorA dimerization or DNA binding, and therefore the arrangement of the complex is not affected. All protein residues could be built into the electron density apart from several N- and C-terminal residues, so that the final model contains residues 2-95 of chain A and 3-97 of chain B. The structure shows good stereochemistry according to *PROCHECK* (Laskowski *et al.* 1993) and *WHATCHECK* (Hooft *et al.* 1996) and contains one KorA (chains A 2-95 and B 3-97) dimer, one complete 18-bp O_A* duplex with a two-fold disorder, and 317 solvent molecules.

	Native-like merged peak
Resolution range (Å)	50-1.85
R_{work} (%)	17.0
R_{free} (%)	20.1
Reflections in R_{work}	32,666
Reflections in R_{free}	1,753
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.522
No. of atoms	
Protein and DNA	2961
Water oxygens	317
Ramachandran (%)	
Most favored	98.8
Additional allowed	1.2

Table 4.3: Refinement statistics.

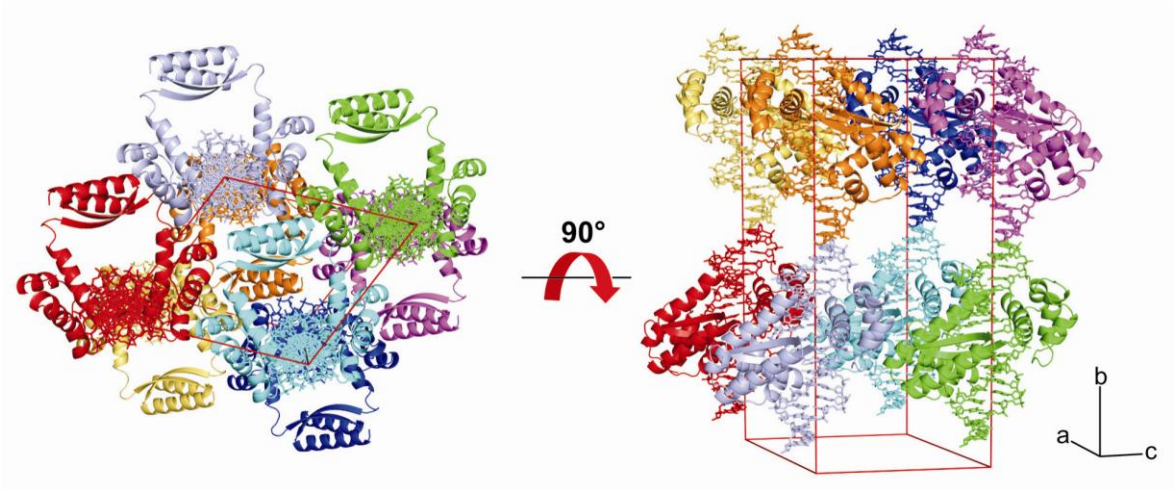


Figure 4.11: Crystal packing of the KorA-O_A* complex. This figure shows the unit cell in two orientations. The unit cell consists of two asymmetric units each containing one protein-DNA complex (space group $P2_1$). Each complex is visualized in a different color, with the DNA duplexes stacked pseudo-continuously on top of each other along the crystallographic *b* axis. For clarity, the DNA is shown in only one of the two orientations present in the complex. Unless stated otherwise, pictures were prepared using *PyMOL* (DeLano 2002).

4.2 The structure of the KorA-O_A* complex

4.2.1 The KorA monomer

The KorA monomer is composed of two distinct domains connected by a linker consisting of residues 66-69. A four-helical assembly (α1, residues 7-14; α2, 21-31; α3, 37-44; α4, 48-65) forms the N-terminal DNA-binding unit. The predicted HTH motif is comprised of α3, the scaffold helix, and α4, the recognition helix. On one side, the recognition helix is flanked by the three short helices, α1-α3, oriented up and down perpendicular to α4. One right-twisted β-strand (β1, 72-78) and the adjacent C-terminal helix (α5, 80-97) form the dimerization module. The helix α5 is positioned antiparallel to β1 resulting in a gripper-like shape (Figure 4.12).

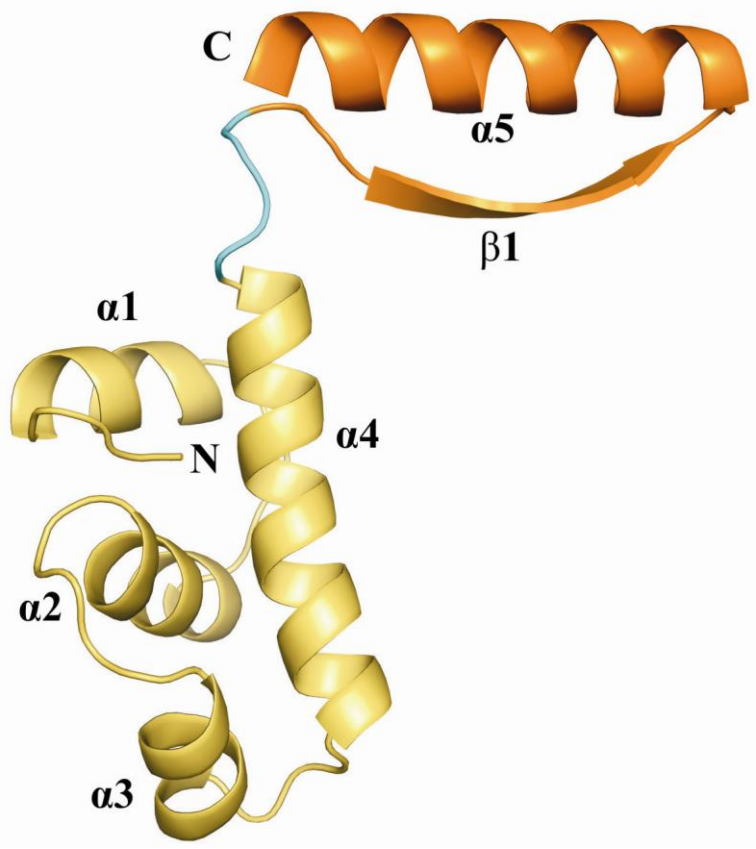


Figure 4.12: The KorA monomer. Schematic representation of the KorA protein with the N-terminal DNA-binding domain (yellow), the linker (cyan) and the C-terminal dimerization module (orange).

4.2.2 The KorA dimer

With their C-terminal gripper the KorA monomers interlock into a very tight homodimer. The β-strands associate to form an antiparallel β-ladder, and the C-terminal α-helices form a two-helical bundle with an interaxial angle of approximately 20° (Figure 4.13). In this arrangement, the helices stretch out along the curved surface of the β-sheet, and residues within the entire length of the α5 helices interact with residues of the β-sheet. Interactions forming the dimer interface initiate in the linker region at residue Leu67 and continue towards the C-terminus of each monomer. The surface area buried in the dimerization interface covers 14.7% (2,115 Å^2) of the total solvent-accessible surface area, ASA, (14,345 Å^2) for both KorA monomers, calculated using the *PISA* server with a probe radius of 1.4 Å (Krissinel and Henrick 2007).

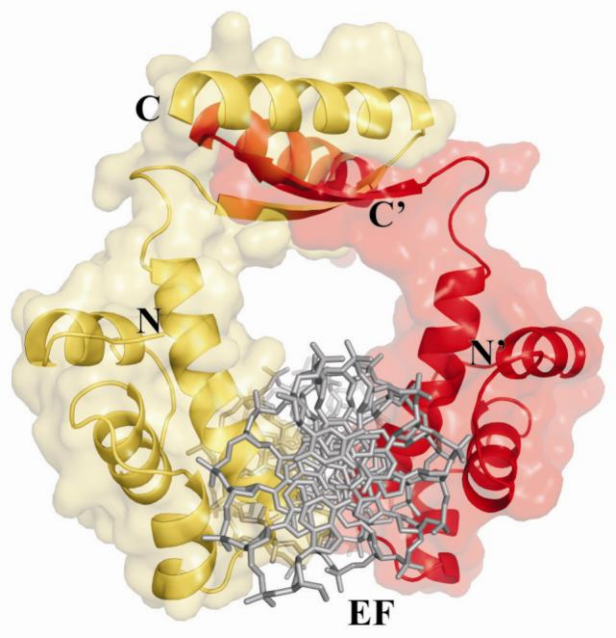


Figure 4.13: The KorA dimer. Image of the KorA dimer (yellow and red) on a semitransparent molecular surface bound to its operator site, looking down the DNA axis. For clarity, the DNA (gray) is shown in only one of the two orientations present in the complex.

Across the interface, side chain interactions are hydrophobic in the core region with the tightly packed aliphatic side chains of Val74, Leu78 and Val86 (see Figure 4.14). The strongest van der Waals interactions (ΔASA > 40 Å^2) arise from residues Tyr71-Pro79 of the β1-strand, and the α5 residues Gln82, and Val86-Glu90, symmetrically contributed by each monomer. The β-strands are connected to an antiparallel β-ladder by eight hydrogen

bonds formed between opposite main chain atoms (Ala72, Val74, Ala76 and Leu78) and the hydrogen bonded side chains of Thr75 directly in the center of the sheet. Hydrogen bonding between backbone atoms of Gly70 and Glu80 positions the N-terminus of helices α5 and α5' on top of the N-terminus of the strands β1' and β1, respectively.

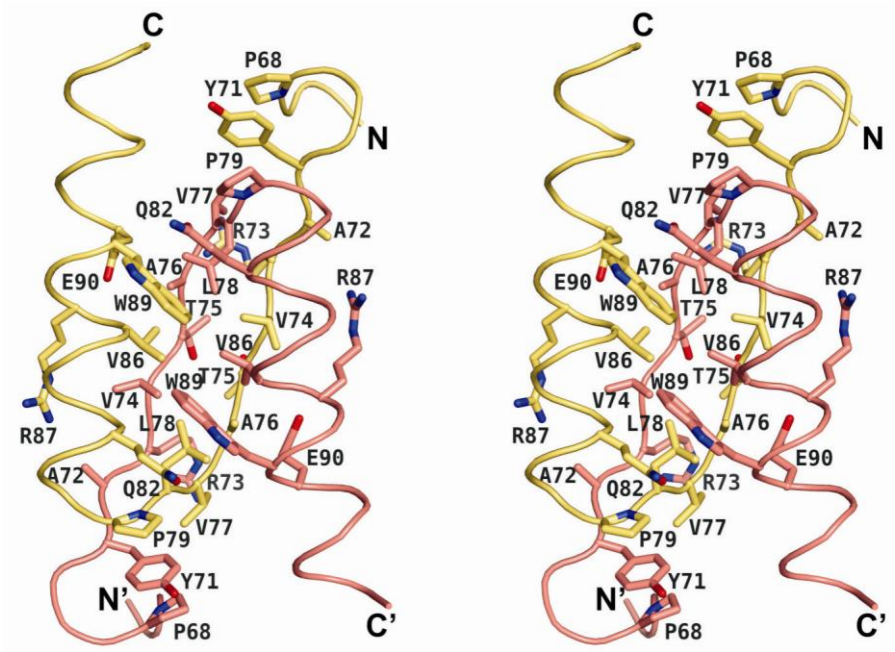


Figure 4.14: Interactions within the dimerization module. Stereo image of the KorA dimerization interface. All residues involved in dimerization are shown for both monomers (yellow and red).

In the crystal structure the dimerization module of the KorA dimer is displaced from the DNA-binding domain, bound to its operator O_A*. The C-terminal domain is elevated above the DNA-binding part of KorA and stabilized through various interactions. First, residues from the linker and dimerization region are involved in stacking-like van der Waals interactions. Side chains of Pro68-A, Tyr71-A and Pro79-B are arranged on top of each other in a sandwich formation with Tyr71-A positioned in the center (Figure 4.14). This arrangement is continued with stacking-like van der Waals interactions of Gln82-B, contacting Pro79-B, and Trp89-A. Due to the two-fold pseudo-symmetry of dimeric KorA, these interactions occur twice in the interface. Second, the hydrogen bonded side chains of Asp64 and Arg73, each belonging to a different subdomain within one KorA monomer, are stabilizing the assembly. In chain B this interaction is water-mediated (Figure 4.15). The spatial separation of the dimerization subdomain from the major groove of the DNA results in a ring-like appearance of the complex assembly with an inner diameter of ~20 Å (Figure 4.13).

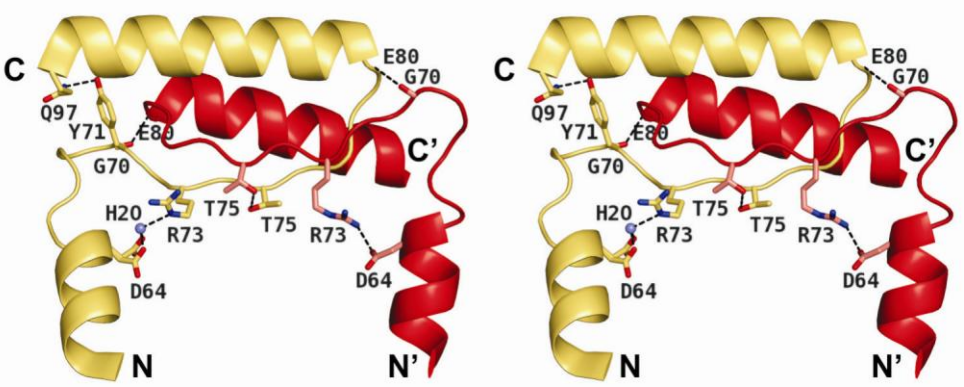


Figure 4.15: Stabilization of the dimerization module. Stereo representation of the KorA dimerization module (yellow and red). Residues involved in stabilization of the ring-like assembly are shown as sticks.

4.2.3 The DNA-binding domain of KorA

In the DNA-binding domain of KorA four compactly folded helices (α1-α4) are arranged around a core of hydrophobic residues, oriented towards the interior of the bundle. Sharp turns between the four helices are mediated by four glycine residues Gly16, Gly20, Gly34 and Gly45. Helices α1 and α3 sandwich α2, all three being positioned on one side of α4. The three N-terminal helices are tethered to α4 by van der Waals interactions, stabilizing the relative orientation of the scaffold and recognition helices (α3 and α4) of the HTH motif. Additional hydrogen-bond interactions, found in both monomers, derive from Gln15 (α1) contacting Lys65 of α4, from Thr23 (α2) recognizing Arg57 of α4, from Gln37 (α3) binding to the backbone of Ser52 (α4), and Arg4 connects the N-terminus with with the backbone of residues Leu31 and Val32, both belonging to α2. Hydrogen bonding between Glu7 of α1 and Arg28 of helix α2 is only seen in one monomer (Figure 4.16).

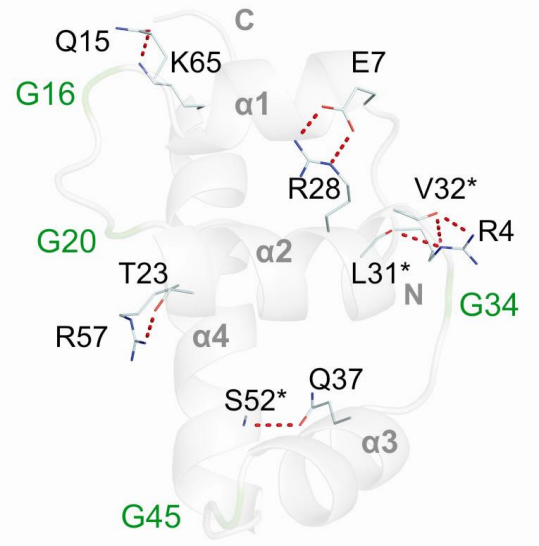


Figure 4.16: The DNA-binding domain. Representation of the DNA-binding domain of one KorA monomer (gray). Residues involved in stabilization of this assembly by forming hydrogen bonds between different helices are shown. Residues that form hydrogen bonds with their backbone are marked with an asterisk. Glycine residues that mediate sharp turns between the helices are shown in green.

4.2.4 Orientation of the operator site

The synthetic 18-bp O_A* DNA duplex is perfectly symmetric within the 12-bp operator site. The three base pairs on both ends outside the consensus sequence lack in symmetry. They contain three brominated deoxyuridine nucleosides at positions 2, 16 and 17 on the anti-sense strand (Figure 4.17 A). Upon analyzing the heavy-atom sub-structure, six bromine sites were detected, permitting to fit the DNA into the electron density in two alternative orientations. Subsequent refinement of the complex with only one orientation present resulted in significant difference density at the terminal base pairs (Figure 4.17 B). Thus, the DNA adopts a two-fold disorder with approximately 50% occupancy for each orientation (Figure 4.17 C). The structures of the two oligonucleotides (strands EF and GH) superimpose well with an r.m.s. deviation of 1.068 Å for all atoms, as calculated with *LSQKAB* (Kabsch 1978). Within the symmetric 12-bp operator site the r.m.s. deviation is reduced to 0.659 Å for all atoms. In the crystal, the 18-bp oligonucleotides are stacked with a negative twist in the junction along the crystallographic *b* axis, giving rise to a pseudo-continuous double helix, shown previously in Figure 4.11.

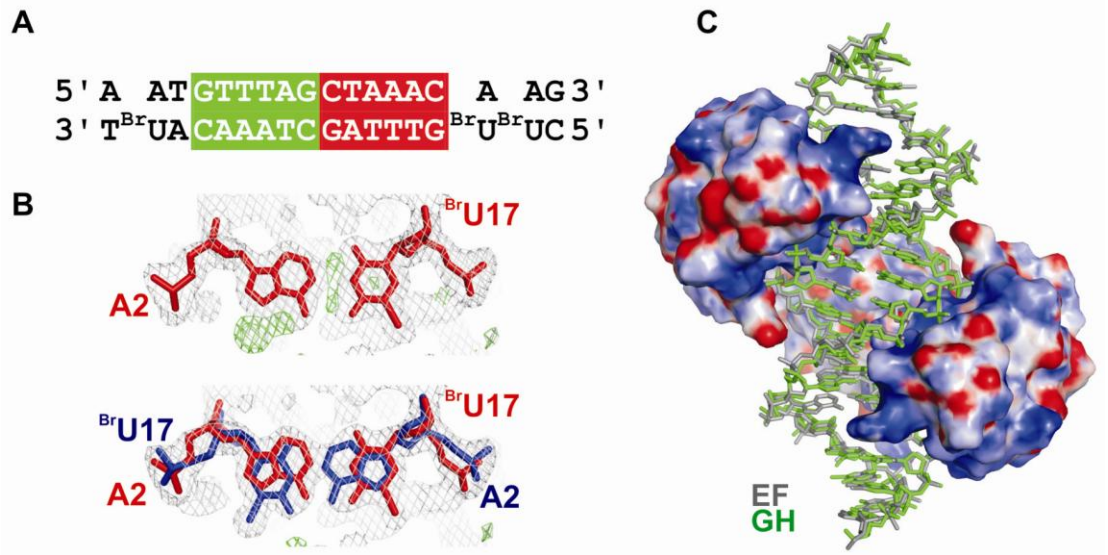


Figure 4.17: Orientation of the operator binding site. (**A**) O_A* with the central 12-bp consensus operator sequence in red and green. (In O_A the ^{Br}U bases were replaced by T.) (**B**) Two-fold disorder of KorA-bound O_A* DNA. Top, base pair A2-$^{Br}U17$ with electron density before allowing for two-fold disorder of the KorA-bound DNA. Bottom, A2-$^{Br}U17$ base pairs of the superimposed DNA duplexes EF (red) and GH (blue) after refinement. The $2F_o$-F_c density (gray) is contoured at 1.0 σ, the F_o-F_c difference density (green) is contoured at 3.0 σ. (**C**) Electrostatic surface potential of KorA; positive charge of the protein in blue and negative charge in red. The DNA is bound to the KorA dimer in two orientations with duplexes EF (gray) and GH (green).

4.2.5 The operator binding site of KorA

KorA binds the 18mer DNA as a dimer, with each monomer contacting one half-site of the operator fragment. The electrostatic surface potential of the DNA-binding domain reveals a positively charged patch facing the bound DNA (Figure 4.17 C). KorA recognizes the operator sequence specifically through the recognition helix, which is inserted into consecutive major grooves of the operator site, as observed in many prokaryotic HTH-containing proteins. The N-terminus of the recognition helix is imbedded in the major groove making direct base contacts through its side chains. The recognition helix is oriented perpendicular to the helical axis of the DNA and profoundly penetrates the major groove. The DNA is bent so as to allow full access of KorA side chains into the adjacent major grooves. The position of the recognition helix is stabilized by hydrogen bonds between residues at the N-termini of helices α1-α3 and the sugar-phosphate backbone of the DNA, sealing the major groove above the contact site of the recognition helix. Contacts to the minor groove of the DNA are not observed.

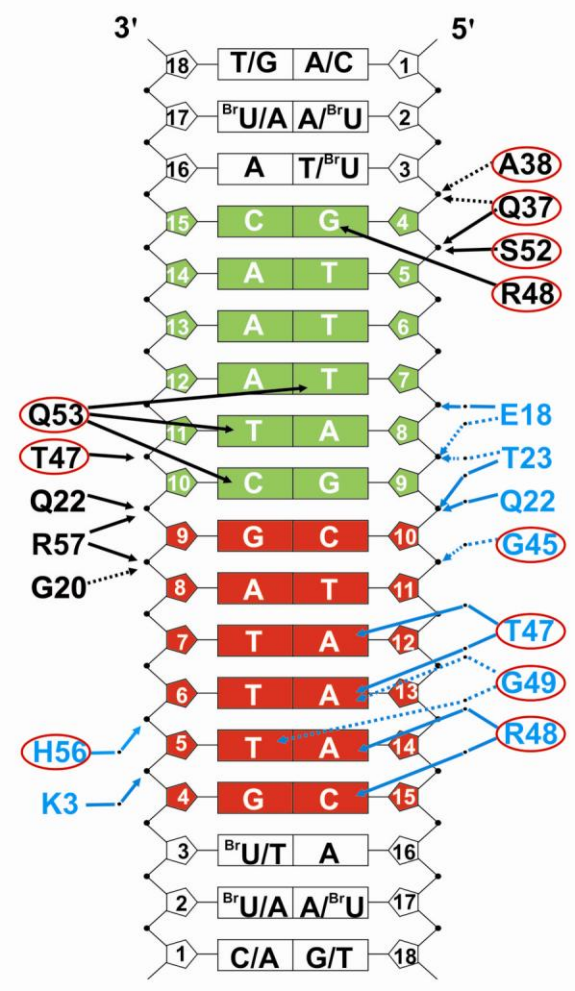


Figure 4.18: The operator binding site. The symmetric operator has two half-sites, each of which binds one KorA monomer with identical geometry. Two different types of protein-DNA interactions are shown in the two half-sites. In the top half-site (green), only direct protein-DNA contacts are shown, whereas water-mediated interactions are shown for the bottom half-site. Both types of interactions occur simultaneously in both half-sites of the operator DNA. Amino acids circled in red are from the HTH motif of KorA. Direct hydrogen-bonded contacts involving protein side chains and DNA are indicated by black solid arrows, those between protein backbone and DNA by black dotted arrows. Contacts mediated by water molecules are marked by blue arrows, those between protein side chains and DNA by blue solid lines, and those between protein backbone and DNA by blue dotted lines.

In the crystal structure each KorA monomer engages in a total of 27 identical contacts, of which 13 are direct and 14 are water-mediated (Figure 4.18). The two-fold disorder of the DNA does not affect protein-DNA interactions, because KorA contacts only the perfectly palindromic 12-bp operator sequence and not the flanking three terminal base pairs, which lack in symmetry.

4.2.9 The conformation of the operator site

CURVES 5.3 analysis (Lavery and Sklenar 1988) of the pseudo-symmetric DNA duplexes shows that the DNA in the crystal is smoothly bent by 14° and adopts, essentially, a standard B-DNA form. The mean helix twist in the two duplexes is 33.3°, with a rise of 3.3 Å per base-pair step. The major grooves are widened from 11 Å to 15 Å upon insertion of the recognition helix. The binding of the Arg48 side chain to the G4 base is associated with a local overwinding of the DNA helix at the G4-T5/A14-C15 and the T5-T6/A13-A14 base-pair steps (twist values of 36° and 42°, respectively). Similarly, the Gln53 binding to bases T7, C10 and T11 is linked with a slight overwinding of the helix at these base-pair steps (see Figure 4.18).

5 Discussion

5.1 Success and failure of experimental MAD phasing

Measuring the polarized absorption spectrum of the protein-DNA crystal in order to determine the optimal wavelengths for MAD phasing did not result in interpretable scattering curves (Figure 4.10). Only the collection of additional data sets at “possible” inflection point wavelengths allowed to overcome the phase problem. After the successful structure analysis of the KorA-O_A* complex, the polarization-dependence of anomalous scattering in brominated DNA molecules was described in detail (Sanishvili *et al.* 2007; Schiltz and Bricogne 2008).

Chemical bonding and the symmetry of the heavy atom’s local environment result in a directional dependence of f' and f'' on the direction of linear polarization of the X-ray beam, called anisotropy of anomalous scattering (Hendrickson *et al.* 1988; Schiltz *et al.* 2004; Dmitrienko *et al.* 2005). In brominated DNA molecules the heavy atom is bound to the base aromatic ring system (Correll *et al.* 1997) giving rise to significant anisotropy of anomalous scattering at the Br *K* edge (Templeton and Templeton 1995; Schiltz *et al.* 2004).

Synchrotron sources generate linearly polarized X-radiation. In a MAD phasing experiment, the correct choice of crystal orientation with respect to the direction of the synchrotron radiation polarization can be as crucial for the quality of phases as is the correct choice of the wavelength. The f'' spectra show substantial variations with crystal orientation, as seen in Figure 5.1 for a 5-bromouracil crystal. The magnitudes of f' and f'' are maximal in the direction along the C-Br bond.

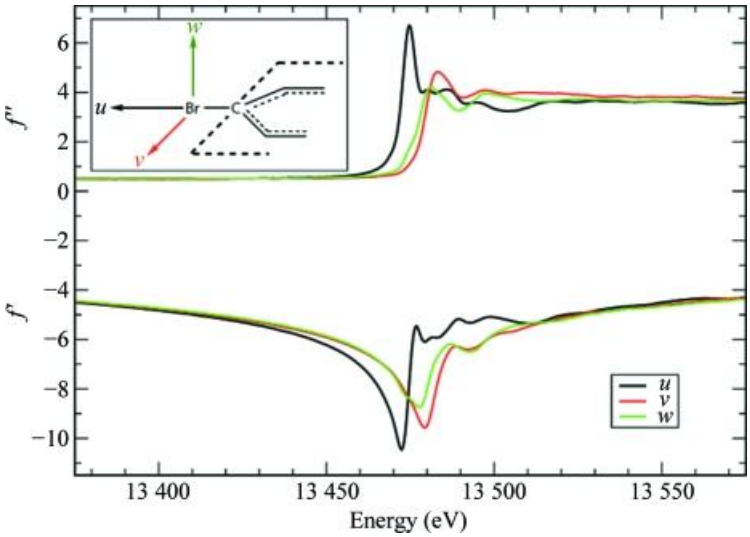


Figure 5.1: Polarization-dependence of anomalous scattering. The anomalous scattering curves of brominated nucleotides are shown for the different polarization directions of the incident X-ray beam; *u*, *v* and *w* in black, red and green, respectively. Data from Sanishvili *et al.* (2007).

In brominated nucleic acid structures (B-DNA or Z-DNA), the packing of the molecules in the crystal is such that all C-Br bonds in the unit cell are located in planes almost perpendicular to the helix axis, but they are not parallel to each other within the planes. This results in a net overall directional dependence of anomalous scattering. Thus, the orientation of the crystal results in enormous energy variations of the extremum values in the f' and f'' spectra. In the KorA-O_A* complex the recognition helix of KorA is deeply buried in the major groove, and forces the DNA to bend significantly to allow full access of KorA side chains into the adjacent major grooves. As a result, the stacking of bases within one molecule results in a considerable deviation of the orientation of all C-Br bonds, as seen in Figure 5.2. However, *CURVES 5.3* analysis (Lavery and Sklenar 1988) confirmed that O_A* adopts, essentially, a standard B-DNA form. Thus the net overall directional dependence of anomalous scattering might be comparable to perfect B-DNA.

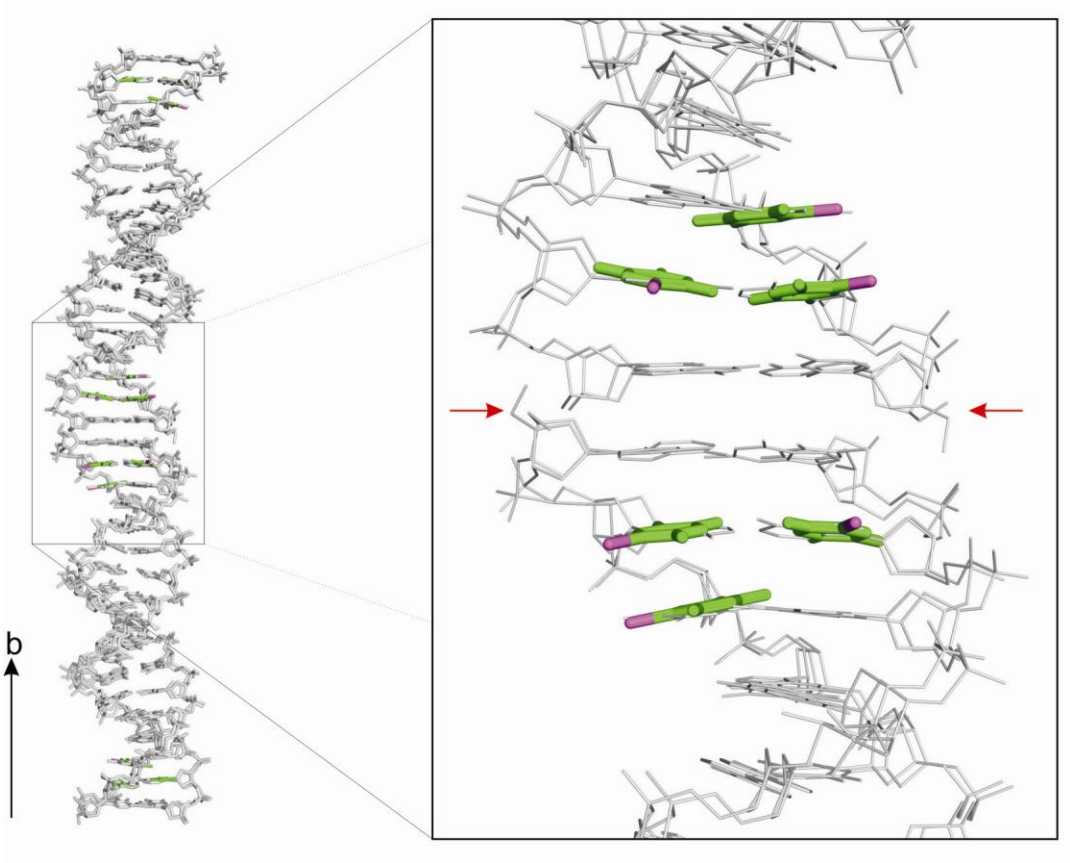


Figure 5.2: Crystal packing of the brominated O_A* oligonucleotides. The two orientations of the pseudo-continuous 18-bp operator DNA are demonstrated in gray. The ^{Br}U moieties are displayed in green, with the Br atoms highlighted as magenta sticks. The C-Br bonds are approximately coplanar, almost perpendicular to the helix axis which is along the crystallographic *b* axis, and the bond vectors point in opposite directions. The close-up on the right side demonstrates the intersection of the pseudo-continuously stacked operator DNAs, marked with red arrows.

The recent results about the polarization-dependence of anomalous scattering provide a possible explanation for the observations of the fluorescence scan. During fluorescence scanning, the angular position of the KorA-O_A* crystal to the polarization direction of the X-ray beam was most likely not in the optimal range. The success of this experiment was therefore entirely based on guessing the correct inflection point wavelength. After collecting three different inflection point data sets, the radiation damage of the crystal was too severe to collect a remote data set. In the merged peak data set, residual effects of anisotropic anomalous signals remain present, which probably resulted in artificially low occupancy values for the bromine atoms (0.25 in the two-fold DNA) during refinement. It cannot be excluded, however, that a fraction of the label was lost due to debromination of the 5-bromodeoxyuridines either before or during X-ray data collection. The observed fractional occupancy of the bromine sites thus is likely the result of a combination of the anisotropy of anomalous scattering and debromination.

5.2 Specific DNA-binding of KorA

Specific DNA binding of KorA to its operator site was studied by various methods. First, it was confirmed by gel filtration and analytical ultracentrifugation that KorA is a very stable homodimeric protein. The full-length transcriptional repressor KorA (residues 1-101) bound to the pseudo-symmetric 18-bp oligonucleotide, O_A*, of sequence 5'-$C^{Br}U^{Br}$U GTT TAG CTA AAC A^{Br}UT-3', was then crystallized to analyze this protein-DNA complex assembly in detail. Outside of the 12-bp operator sequence, three 5-bromo-deoxyuridine nucleosides were introduced into one DNA strand to allow MAD phasing. Placed within the flanking base pairs outside the operator binding sequence, 5-bromouracil was not expected to interfere with specific protein-DNA interaction. The complex formation between KorA and 18-bp DNA was not disturbed when using the brominated oligo-nucleotide, as determined by native PAGE and gel filtration experiments.

Homodimeric repressor proteins, like KorA, are known to recognize symmetric operator sequences. Binding of the DNA can therefore occur in two possible orientations. Taking this into account, it was not surprising to observe a two-fold disorder of the oligonucleotide in the crystal structure. Within the 12-bp operator sequence the two duplexes super-impose perfectly, allowing the two KorA monomers to form identical contacts. The flanking three terminal base pairs, which lack in symmetry, were not contacted by KorA in the crystal structure.

The KorA-O_A* complex shows remarkable structural complementarity between the repressor's DNA-binding domain and the 12-bp operator binding site. KorA binds the

palindromic DNA symmetrically with its HTH motif. Each monomer contacts the major groove on one half-site of O_A. The specificity of KorA for its O_A site is primarily based on two side-chain interactions from the recognition helix (α4) reaching deep into the major groove to make several base contacts (Figure 5.3 A, Table 5.1). The Arg48 side chain binds to the O^6 and N^7 atoms of G4, and Gln53 contacts the three bases T7 (O^4), C10 (N^4) and T11 (O^4). Each specific contact is made to the up and to the down oriented DNA duplex with very similar hydrogen bond lengths (Table 5.1).

The importance of specific contacts between KorA and the operator O_A is supported by electrophoretic mobility shift assays. Four KorA variants were designed. In the first two mutants only one of the two specific DNA-binding residues was changed to alanine (R48A and Q53A), in the third both specific binders (R48A/Q53A) were deleted. According to the binding mode observed in the crystal structure (Figure 5.3 A) truncation of the Arg48 side chain should remove two hydrogen bonds per operator half-site, and the Q53A mutation should lead to the loss of three hydrogen bonds per half-site. In addition, Gln53 was changed to the isosteric Glu53, causing the loss of two out of three hydrogen bonds.

Wt KorA and mutants were purified and tested for binding to a 436 bp DNA duplex containing the class I O_A site of *korAp* (Figure 5.3 B). Wt KorA associates specifically with O_A, and no retardation was observed with the nonspecific control DNA fragments that did not carry an O_A site. The introduction of the Q53A, R48A/Q53A and Q53E mutations completely abolished the ability of KorA to bind specifically to O_A in all cases. R48A retained a reduced capability to shift the operator fragment and showed some nonspecific binding in addition. Taken together, the gel-shift assays strongly suggest that the binding mode observed in the crystal reflects the mode of operator recognition by KorA in solution, and they identify Gln53 as the crucial residue for specific operator binding.

To validate the strong binding capacity of wt KorA, ITC binding studies were performed using the non-brominated 18mer duplex O_A containing the class I operator sequence. ITC data yielded K_D values of 23.3 ± 0.7 nM for KorA:DNA (2:1) binding at 37 °C, as shown in Figure 5.3 C. Hence, it was demonstrated that one KorA dimer binds exactly one operator binding site with strong affinity.

These results finally explain why KorA binds two classes of operator sites on the RP4 plasmid with different affinities (see Table 1.2). Class I operators (*korAp*, *klaAp* and *trfAp*) are perfect in symmetry with K_D values of 13 to 20 nM while class II operators (*klcAp*, *kfrAp*, *kleAp* and *kleCp*), with more then ten-fold higher K_D values, are lacking either G4 (contacted by Arg48) or T7 (contacted by Gln53) in one half-site of the consensus sequence. These variations in operator sequences provide a powerful way of control. The activity of each transcription factor binding site can be balanced against that of others.

However, this is only a static way of transcription regulation that cannot be modulated upon environmental changes. Adaptable regulation depends entirely on regulation of transcription factors.

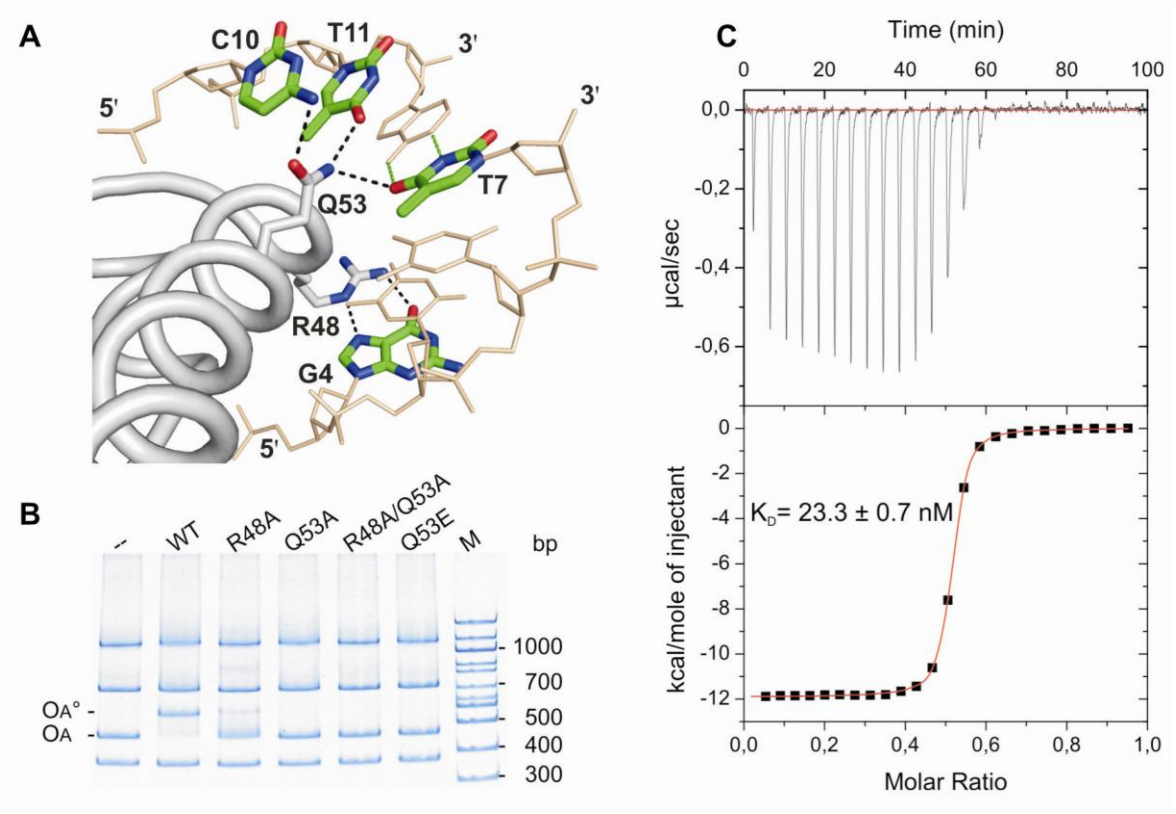


Figure 5.3: Specific binding of KorA to its operator binding site. (**A**) Specific KorA-O_A* contacts in the crystal. Arg48 and Gln53 of the recognition helix (α4) of one KorA monomer form specific hydrogen bonds to G4, T7, C10, and T11, in one half-site of the symmetric operator sequence. (**B**) *In vitro* binding assay of purified wt KorA and mutants binding to DNA. The 436-bp fragment carries the class I O_A site (*korAp*) of plasmid RP4, the other fragments serve as competitor DNA. 7.5 pmol of wt and each mutant protein were applied. The DNA mixture contained 0.15 pmol of each fragment. O_A, DNA fragment containing O_A; O_A°, complex of O_A with wt KorA; M, 100-bp DNA ladder. (**C**) Calorimetric titration of KorA with 18-bp O_A. The panel shows the integrated heat released after correction for dilution (data points, squares) and the curve of best fit (red) for one KorA dimer binding to a single O_A site.

KorA	**O_A***	**Bond lengths (Å)[a]**
R48 ($N^{\eta 2}$)	G4 (O^6)	3.11 ± 0.15
R48 (N^{ε})	G4 (N^7)	3.08 ± 0.11
Q53 ($O^{\varepsilon 1}$)	C10 (N^4)	3.07 ± 0.31
Q53 ($N^{\varepsilon 2}$)	T11 (O^4)	2.92 ± 0.06
	T7 (O^4)	2.90 ± 0.17

[a]Calculated from four different bond lengths deriving from contacts of the symmetric KorA dimer to the two-fold disordered 18-bp O_A* DNA.

Table 5.1: Specific binding of Arg48 and Gln53 to O_A.

5.3 Structural homologs of KorA

Structural similarity searches using *DALI* (Holm and Sander, 1995) found no homolog for the full-length KorA. When the search was limited to the DNA-binding unit (residues 2-65) a close resemblance to the YlxM/p13-like family could be found, such as the hypothetical UPF0122 transcription factor SAV1236 from *Staphylococcus aureus* (PDB entry 1xsv) with a DALI Z-score of 7.7 and an r.m.s. deviation of 2.2 Å (Figure 5.4 A). With the search focused on the DNA recognition region, the orientation of the scaffold and the recognition helix reveal structural homology to the C-terminal σ_4 domain of FliA (σ^{28}) from *Aquifex aeolicus* (1rp3) and to the C-terminal HTH motif of the GerE-like TraR protein from *Agrobacterium tumefaciens* (1h0m) with Z-scores of 4.1 and 3.6, and r.m.s. deviations of 1.3 Å and 1.8 Å, respectively (Figure 4.16 B). Sigma factors and members of the GerE-like family are key regulators of bacterial transcription; σ_4 specifically recognizes the -35 promoter element with its HTH motif (Vannini *et al.* 2002; Sorenson *et al.* 2004). This result shows that the full-length structures of DNA-binding proteins can be very diverse. The recurring HTH motif is presented in exclusive ways by each protein to achieve specific sequence recognition.

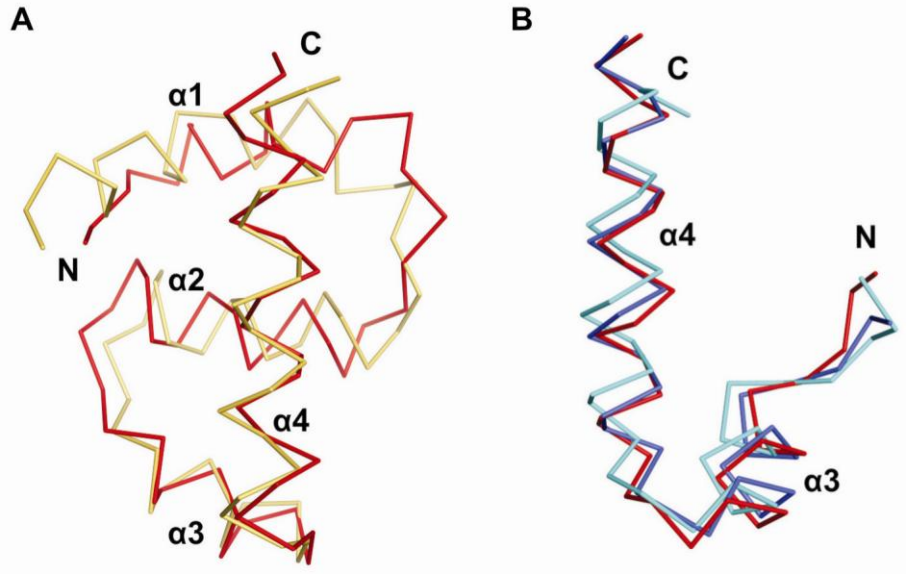


Figure 5.4: Structural comparison. (**A**) Superposition of KorA DNA-binding domain (red) and SAV1236 (yellow). (**B**) Superposition of α3 - α4 of KorA (red) with FliA (blue) and TraR (cyan).

5.4 KorA homologs of other IncP-1 plasmids

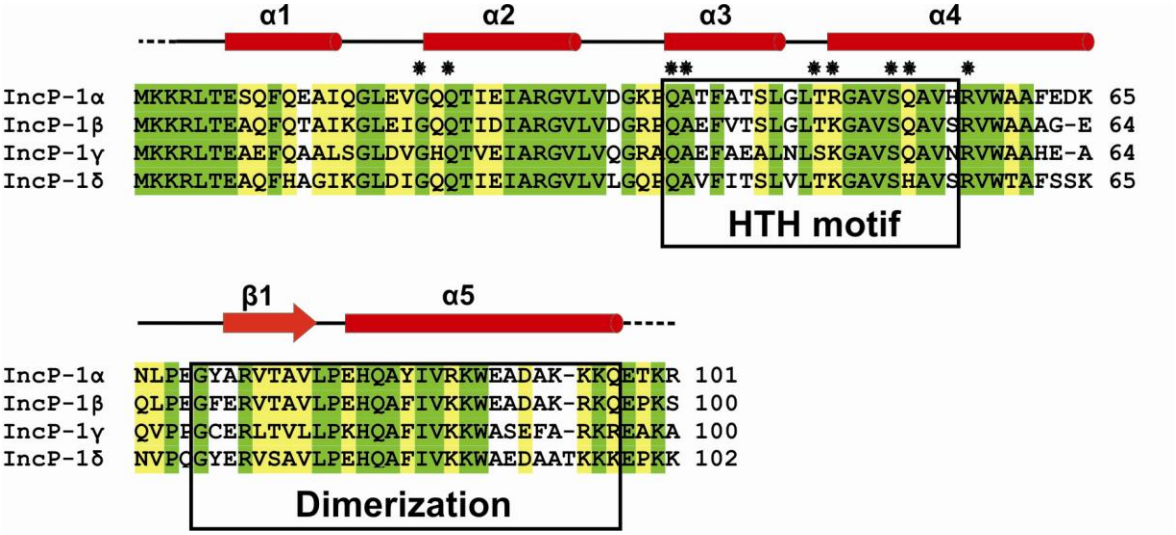


Figure 5.5: KorA homologs from IncP-1 plasmids. Multiple sequence alignment of KorA proteins of the IncP-1 plasmids RP4 (α), R751 (β), pQKH54 (γ) and pEST4011 (δ) with accession numbers and shared identity towards full-length $KorA_{RP4}$: P03052, 100%; Q57423, 79%; Q1H9Z5, 60%; Q6QHR0, 69%. The secondary structure elements of KorA are represented above the aligned sequences. Identical and similar residues are highlighted in green and yellow, respectively. Dots mark KorA residues specifically involved in DNA interaction. Sequences corresponding to the HTH or dimerization motifs are enclosed in boxes.

IncP-1 plasmids fall into the α, β, γ, and δ subgroups (Dennis 2005; Bahl *et al.* 2007). The $KorA_{RP4}$ and $KorA_{R751}$ proteins (Figure 5.5; sequence identity 79%) belong to the best studied IncP-1 plasmids, and recognize exactly the same operator sequence but prefer binding to their own plasmid-specific O_A sites which only differ in the flanking DNA sequences (Table 5.2). All residues of $KorA_{RP4}$ found to participate in direct protein-DNA contacts in the crystal structure (Figure 4.18) are almost identical in both proteins. One exception is Arg48, shown above to be important for operator recognition, which is a lysine residue in $KorA_{R751}$. In addition, His56, involved in water-mediated DNA-interactions, is a serine in $KorA_{R751}$. Mutagenesis of His56 to serine in $KorA_{RP4}$ leads to an increased affinity for all operators, especially for $O_{A/R751}$ (Kostelidou and Thomas 2002). Considering the high sequence similarity, it is most likely, that both KorA proteins exhibit the same fold. Replacing Arg48 by a long and flexible lysine residue would probably result in specific recognition of the G4 base within the O_A site as well, but one hydrogen bond will be lost. There is no simple model describing the one-to-one correspondence between amino acid side chains and DNA bases they contact. Many side chains contact more than one type of base, and each base can interact with different side chains. It seems unlikely, however, that Ser56 contacts the DNA phosphate backbone in the same way as His56. Therefore, the DNA needs to be bent more than seen in the $KorA_{RP4}$ crystal structure, and

maybe the flanking sequences of the operator sites, AT-rich in RP4 and GC-rich in R751, are important for the extent of DNA-bending (Table 5.2). It remains noteworthy that both KorA repressors function only suboptimally on their own plasmids, but equally strong if compared to each other. In transcription regulation operators that function suboptimal are advantageous, because they are amenable to upregulation when necessary.

Operator	Promoter	RP4[a]	R751[a]
O_A1	*korAp*	AAT **GTT TAG CTA AAC** TTC	ATG **GTT TAG CTA AAC** TTC
O_A2	*kfrAp*	ACT **GTT CAG CTA AAC** TCT	AAA **GTT CAG CTA AAC** CGC
O_A3	*trfAp*	AAT **GTT TAG CTA AAC** TAG	CAG **GTT TAG CTA AAC** TGA
O_A4	*klcAp*	ATA **GTT TAG CTA AAT** TTG	TAG **CTT TAG CTA AAG** TGC
O_A5	*kleAp*	ATT **GTT TAG CTA AAT** TTC	ATA **GTT TAG CTA AAG** GTG
O_A6	*kleCp*	ATT **GTT TAG CTA AAA** TTG	not present
O_A7	*klaAp*	AAA **GTT TAG CTA AAC** TTC	not present

[a]The 12-bp consensus sequence is shown in red, base mismatches are highlighted in gray.

Table 5.2: Conservation of KorA operators between RP4 and R751 (Thorsted *et al.* 1998).

5.5 The dimerization module

The crystal structure of KorA provides for the first time the architecture of this unique dimerization module indicating very stable protein-protein interactions. This domain exhibits a high level of conservation between KorA and two other RP4 proteins, TrbA and KlcB. Considering the high sequence similarity, it is most likely, that all three proteins exhibit the same fold within this domain. It could be shown in the past that KorA and KlcB are able form heterodimers *in vitro*, and *klcB* induction was able to release KorA repression of the *kleA* promoter (Bhattacharyya and Figurski 2001). However, it is still unclear if this is the actual case under physiological levels of both proteins. If so, it offers the possibility of transcription factor activity regulation. The DNA-binding affinity of KorA may be modulated by binding to KlcB depending on protein concentration.

The DNA-binding and the dimerization domain of KorA are each very compact and stable, whereas the linking region shows potential flexibility. In gel filtration experiments, it was established that the KorA dimer does not retain a globular fold, and therefore elutes earlier as expected from its molecular weight. As seen for many homodimeric transcription regulators, this flexibility is probably the reason why KorA could never be crystallized in the unbound state. Complex formation with DNA results in a much higher stability of the KorA dimer.

A very interesting feature revealed by the structural analysis of the complex is the ring-like conformation. In solution the conformations of KorA and O_A* in their unbound and complexed states are not necessarily rigid, being in a dynamic equilibrium with their surroundings. The ring-like assembly of the KorA-DNA complex may be expected to be deformable, and a flexible attachment of the C-terminal domain to the DNA-bound core of the protein is most likely biologically relevant. KorA is known to cooperate with another major RP4 repressor protein, KorB, through its C-terminal domain from various distances (Kostelidou *et al.* 1998). For these direct interactions the ring-like assembly could potentially serve as an adaptable anchor site.

To understand the interaction between KorA and KorB in the transcriptional regulation of RP4 genes, the sequences of the RP4 and R751 proteins KorA, TrbA and KlcB, all known to cooperate with KorB (Kostelidou *et al.* 1999), were aligned using CLUSTALW (Thompson *et al.* 1994). KorA and TrbA are global repressors in these plasmids, while the function of KlcB is still unknown (Bhattacharyya and Figurski 2001). Over the length of the matched sequences, a variable degree of sequence conservation is observed (Figure 5.6 A) indicating similar three-dimensional structures. The repressor proteins KorA and TrbA share an HTH motif. Sequence conservation within this motif is limited, reflecting the different operator sequences on the RP4 or R751 genomes to which the repressors bind. A strikingly high degree of sequence conservation characterizes the dimerization module with identical residues at 10 sequence positions between Pro68 and Lys96 of $KorA_{RP4}$. For the KorA and TrbA proteins this conservation is found at the C-terminus, while $KlcB_{RP4}$ and $KlcB_{R751}$ include additional 116 or 154 residues, respectively, adjacent to this region.

Highlighting the preserved region on the exposed molecular surface of KorA reveals a large patch of identical and type-conserved amino acids at the top of the dimer region (Figure 5.5 B). Positioned directly in the center are the exposed hydrophobic side chains of Ile85 and the polar aromatic residues Trp89, His81 and Tyr84 of each monomer (Figure 5.5 C). All four residues are known to favor protein-protein interface locations (Jones and Thornton 1997). Together, they create a surface patch which could serve as the site of interaction with KorB, a finding which is consistent with a recent NMR study that identified Tyr84 as a critical residue for KorA-KorB cooperativity (Bingle *et al.* 2008). It would seem formally possible that, alternatively, protein interactions would involve the opposite, inner side of the dimerization region which faces the DNA. In KorA, the distance between this inner surface and the DNA helix is large enough to permit further inter-molecular interactions, also possible for TrbA, which seems to include a longer linker region, resulting in a considerably increased diameter of its ring-like assembly. KlcB has no predicted DNA binding ability and therefore retains no ring-like conformation. However,

in the KorA structure the exposed residues of the inner side of the dimerization module are Arg73 and Thr75 (see Figure 4.15). These residues show no conservation in the alignment (Figure 5.5 A). I conclude, therefore, that the outer surface of the dimerization regions of KorA, TrbA and KlcB serves as contact site for the flexible and structurally unknown N-terminal domain of KorB.

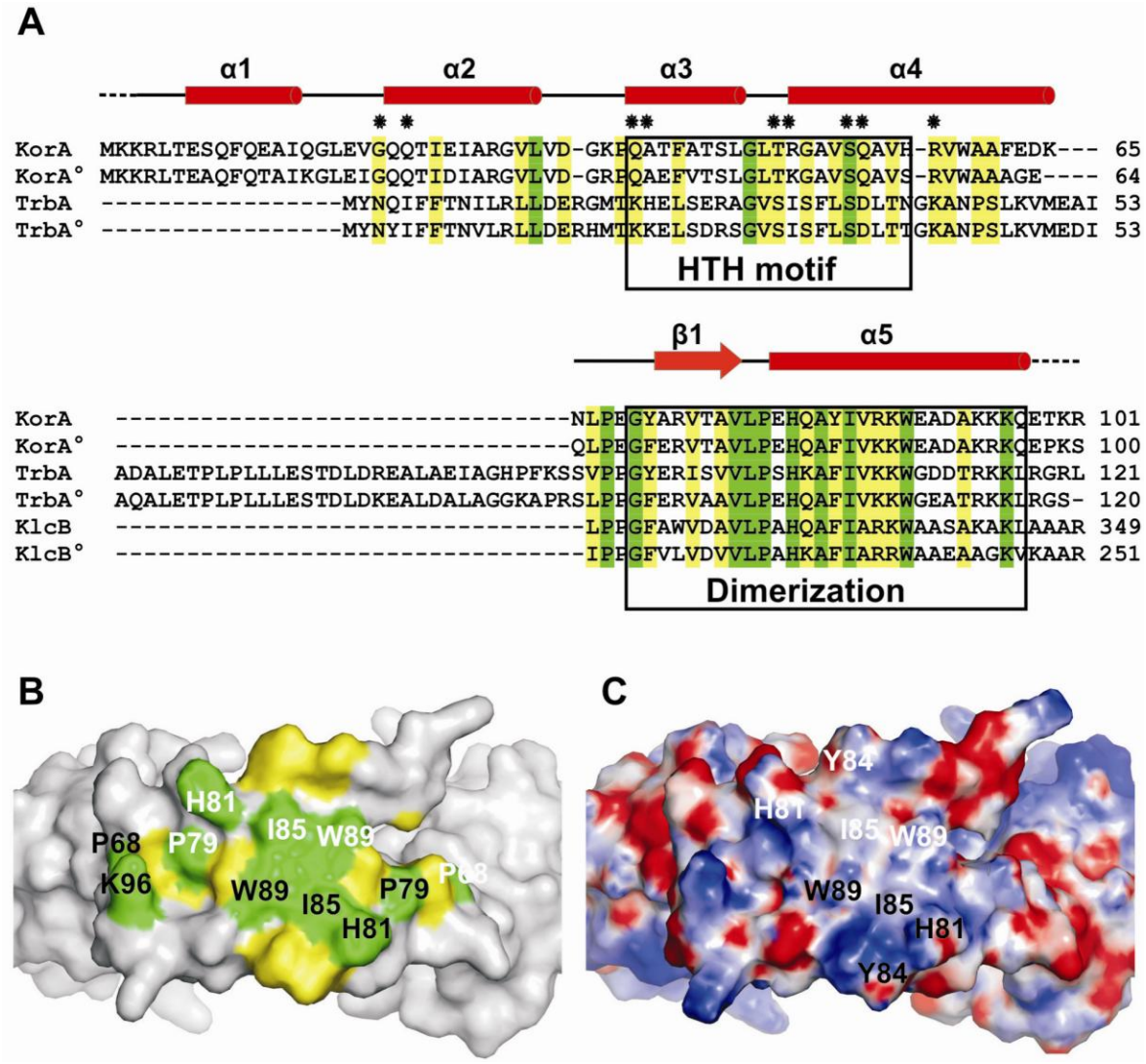


Figure 5.6: The KorA dimerization module. (**A**) Multiple sequence alignment of RP4 and R751 (R751 members marked with °) proteins KorA, TrbA and KlcB (accession numbers as listed above and shared identity towards full-length $KorA_{RP4}$: P03052, 100%; Q57423, 79%; Q01716, 51%; Q56468, 59%; P52605, 32%; and P52604, 41%). The secondary structure elements of KorA are represented above the aligned sequences. Identical and similar residues are highlighted in green and yellow, respectively. Dots mark KorA residues specifically involved in DNA interaction. Sequences corresponding to the HTH or dimerization motifs are enclosed in boxes. (**B**) The molecular surface of the KorA dimerization region (top view) reveals a patch of highly conserved residues. Side chains of identical and similar residues, as seen in the sequence alignment (A), are highlighted in green and yellow, respectively. Residues of one KorA monomer are labeled with the same color (black or white). (**C**) Electrostatic surface potential of the KorA dimerization region (top view); calculated using *APBS* (Baker *et al.* 2001) in a range from -10 kT (red) to +10 kT (blue). Residues possibly involved in protein-protein interactions are labeled in black and white.

5.6 Cooperation mechanism in RP4

The expression of RP4 genes is tightly controlled by the two main regulators KorA and KorB, and it reacts very sensitively to fluctuations in repressor levels. During exponential growth of the host, there are approximately 2000 KorA and 1500 KorB dimers in the cell (Jagura-Burdzy and Thomas 1995; Chiu *et al.* 2008), binding to different classes of operator sites with different affinities. Up to now, the role of second or third class operators is not clearly understood. The suboptimal operator sites grant the opportunity for upregulation whenever suitable conditions emerge; for that reason they only play a supporting role in this control circuit. A crucial role in transcription regulation is played by the cooperation of transcription factors such as KorA and KorB, but the question is how.

It appears conceivable that the driving force for direct interaction is provided by anchoring KorA and KorB in close proximity as bound to DNA. They could then stabilize their association with the DNA through interactions with one another. When bound to DNA, the effective, local concentration of the two proteins at the binding site is much higher than in the bacterial cytoplasm. Thus, even when the proteins might not form stable complexes in solution, interaction between them when bound to DNA would account for cooperativity.

Recent NMR studies could prove the direct interaction of KorA and KorB in solution (Bingle *et al.* 2008). These protein-protein contacts could increase the association rates by helping one another to locate and bind the operator sites more easily, especially at low protein concentrations. This is important, because KorA or KorB alone might arrest transcription of their own genes at the *korAp* site (class I) before any of the low-affinity binding sites are bound. However, interactions between the two global regulatory proteins are not likely to be very tight. First of all, at only five out of the seven RP4 operons that are repressed by KorA an O_B site is also present. Hence, KorA needs to be available as homodimer for the other two O_A sites as well. Second, promoter sites that are regulated by KorA and KorB could lose the possibility of fine-tuning.

The structure of the KorA-O_A* complex provides the first view of a global transcriptional regulator of the RP4 plasmid that cooperates with the repressor protein KorB. This study shows how KorA dimerizes and how the KorA dimer specifically contacts its operator site, suggesting a model for the sophisticated mechanism of repressor cooperation in RP4. On the molecular surface of the KorA dimerization domain a prominent patch of residues favoring protein-protein interactions was identified. These residues are conserved in RP4 repressor proteins that interact with KorB. The crystal structures of the C-terminal dimerization region of KorB, KorB-C, with interlocking SH3-like domains (aa 299-358) (Delbrück *et al.* 2002), and the DNA-binding domain KorB-O (aa 137-252) (Khare *et al.* 2004) are known. The highly flexible N-terminal domain (aa 1-136) could not be crystal-

lized. It appears that KorB contacts KorA *via* its mobile N-terminus which could permit repressor-repressor interactions over variable O_B-O_A distances. The center-to-center distance between the operator sites varies from 32 bp (*korAp*) to 33 bp (*trfAp* and *klaAp*) and 36 bp (*kfrAp*). Assuming standard B-form, as seen in the KorA and KorB operator complexes, also for the linker DNA, the flexible N-terminal domains of KorB and the interacting surface on the KorA dimerization module are located at nearly opposite sides of the DNA helix in *korAp* (Figure 5.7). As the operator sites are moving further apart to 33 or 36 bp in *trfAp* / *klaAp* and *kfrAp*, respectively, the helix twist puts the interacting repressor moieties successively on the same side of the DNA, keeping the effective KorB-KorA separation approximately constant. Given the observed flexibility of the KorB N-terminus, this model explains the cooperation between KorB and KorA at RP4 promoters with variably spaced O_B and O_A sites, and may be a structural template for KorB's interaction with the TrbA repressors and KlcB that share the predicted interaction surface. The reported cooperation between KorA and KorB over a spacing of 92 bp at the *kleA* promoter is not explained by this model and may require looping of the DNA or involve a completely different mode of interaction.

The model shows how cooperative binding of the two transcription factors creates a continuous recognition surface. Thus, the repressors operate as a single unit of regulation. These mechanisms are commonly found in eukaryotes, while in prokaryotes, cooperativity of transcription factors is the exception rather than the rule (Courey 2001).

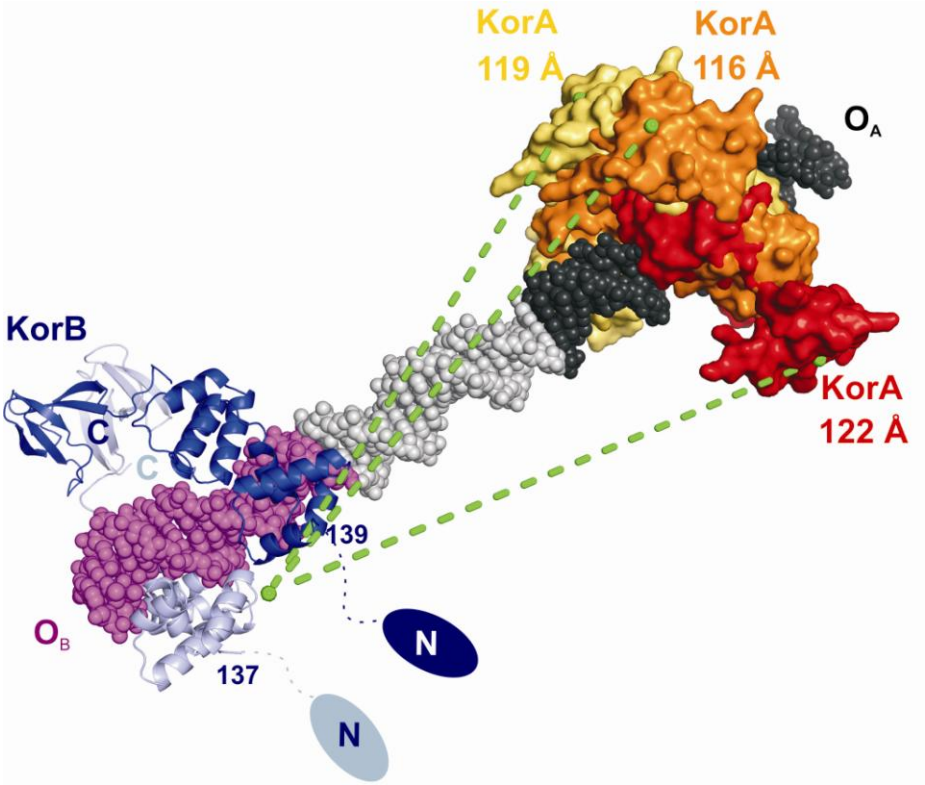


Figure 5.7: Possible interaction between KorA and KorB at RP4 promoters. Crystal structures of the RP4 repressor proteins KorA (yellow, orange and red, each representing the molecular surface of one dimer, respectively) and KorB (chain A, blue, and chain B, light blue; PDB entries 1r71 for DNA-binding domain KorB-O, residues 137-252, and 1igu for the C-terminal domain, residues 299-358) (Delbrück *et al.* 2002; Khare *et al.* 2004) bound to their operator sites. O_A (black) and O_B (magenta) were joined by standard B-form DNA (gray), generated with *MAKE-NA* (http://structure.usc.edu/make-na/) representing the linker region at the *korAp* site. O_A and O_B midpoints are separated by 32 bp (KorA = yellow), by 33 bp (*trfAp* and *klaAp*, KorA = orange), and by 36 bp (*kfrAp*, KorA = red). The structurally unknown N-terminal residues 1-136 of each KorB monomer are symbolized by two ellipsoids in blue and light blue. The distances between the KorB N-terminal region and the KorA dimerization domains in their different settings are indicated by green dashed lines.

5.7 Conclusion and closing remarks

The presented data have made an important contribution to elucidate the cooperation event between the global transcription regulatory proteins KorA and KorB in RP4 plasmids on the molecular basis. The structural characterization of the KorA homodimer bound to its operator DNA has provided a starting point for biochemical analysis of the protein's functions. It was shown that KorA has a high binding affinity towards its operator binding site, and two residues within its HTH motif were identified that are involved in specific DNA binding. In addition, the crystal structure of the KorA-O_A complex has provided valuable insight into the architecture and assembly of a corepressor complex between KorA and KorB, and I could propose a model for gene regulation at O_A/O_B promoters.

It is necessary to understand the molecular processes contributing to cooperativity because they define the basis of the sophisticated gene regulatory circuit in RP4. My results provide a much better understanding of KorA and its role in plasmid maintenance through transcriptional control. The emerging picture, however, leaves many questions unanswered. The task for future work will be to elucidate the interplay of all global repressors in RP4 to get insight into the amazing fine-tuning possibilities of RP4.

6 Summary

Replication and gene expression of plasmids impose a metabolic burden on their host cells. Plasmids have therefore evolved systems to guarantee their own DNA stability and integrity. The stringent regulation of gene expression plays the essential role in the maintenance of plasmids and their hosts. Each step in transcription initiation can be regulated by transcription factors interacting at or near the corresponding promoter, where they either activate or repress the transcription of the gene by interaction with RNA polymerase (RNAP). In eukaryotes, cooperativity between transcription factors is very common to increase activation or repression of a promoter by direct protein-protein interaction or modulation of the local DNA conformation. In prokaryotic transcription regulation only a few examples of cooperativity are found.

KorA is a global repressor in the promiscuous IncP-1α plasmid RP4 which regulates cooperatively the expression of plasmid genes whose products are involved in replication, conjugative transfer, and stable inheritance. The seven operator binding sites of KorA (O_A1 - O_A7) are scattered throughout the RP4 plasmid genome. The aim of this thesis was to determine the high-resolution structure of KorA bound to its operator binding site, and to elucidate important aspects of its molecular mechanisms of gene regulation. The structure of the full-length transcriptional repressor KorA bound to the pseudo-symmetric 18-bp DNA duplex O_A* (5'-$C^{Br}U^{Br}U$ GTT TAG CTA AAC $A^{Br}UT$-3'), that contains the symmetric operator sequence and incorporates 5-bromo-deoxyuridine nucleosides, has been determined by MAD phasing at 1.96-Å resolution. This is the first published crystal structure of the KorA protein.

In the crystal structure, KorA is present as a symmetric dimer and contacts DNA *via* a helix-turn-helix (HTH) motif. Each half-site of the symmetric operator DNA binds one copy of the protein in the major groove. The operator site is smoothly bent and adopts a standard B-DNA conformation. The oligonucleotide shows a two-fold disorder in the crystal structure, however, each specific contact is made to the up and to the down oriented DNA duplex with very similar hydrogen bond lengths. As confirmed by mutagenesis studies, the recognition specificity is based on two KorA side chains, Arg48 and Gln53, forming hydrogen bonds to four bases within each operator half-site. The performed gel-shift assays strongly suggest that the binding mode observed in the crystal reflects the mode of operator recognition by KorA in solution. They identify Gln53 as the crucial residue for specific operator binding.

Analytical ultracentrifugation experiments confirmed the stable dimeric appearance of KorA in solution. Further investigation of the binding mode revealed that KorA binds the operator site as a dimer in a 2:1 molar ratio. The strong binding capacity of wt KorA was validated in ITC binding studies with the non-brominated 18mer duplex O_A (5'-CTT GTT TAG CTA AAC ATT-3') containing the class I operator sequence. Combining structural data and binding assays, these results finally explain how KorA recognizes and binds to two classes of operator sites on the RP4 plasmid with different affinities.

KorA has a unique dimerization module shared by the RP4 proteins TrbA and KlcB. Considering the high sequence similarity of their C-termini, it was presumed, that the KorA dimerization domain might represent the common fold for all. Cooperativity of the global repressor proteins KorA and TrbA with KorB strongly supported that idea. Highlighting the preserved region on the exposed molecular surface of KorA reveals a large patch of identical and type-conserved amino acids at the top of the dimer region. This conserved surface patch led to the conclusion that these proteins cooperate with the global RP4 repressor KorB in a similar manner *via* this dimerization module and thus regulate RP4 inheritance.

7 Zusammenfassung

Replikation und Genexpression von Plasmiden stellen eine metabolische Last für ihre Wirtszellen dar. Infolgedessen haben Plasmide einige Systeme hervorgebracht, die ihre Stabilität und Kopienzahl sichern. Die genaue Regulierung der Genexpression spielt eine entscheidende Rolle bei der Erhaltung von Plasmiden, sowie ihres Wirts. Jeder Schritt innerhalb der Transkriptionsinitiation kann von Transkriptionsfaktoren am entsprechenden Promoter oder in dessen Umgebung beeinflusst werden. Diese Faktoren aktivieren oder unterdrücken die Transkription durch Interaktion mit der RNA-Polymerase (RNAP). In Eukaryoten ist Kooperativität zwischen Transkriptionsfaktoren ein bekannter Mechanismus, um die Aktivierung oder Unterdrückung des Promoters durch direkte Protein-Protein-Wechselwirkungen oder durch Ummodellierung der lokalen DNA-Konformation zu erhöhen. Bei Prokaryoten ist Kooperativität in der Transkriptionsregulierung nur selten vorzufinden.

KorA ist ein globales Repressorprotein des promiskuitiven IncP-1α Plasmids RP4, welches die Expression von Plasmidgenen, die in Replikation, konjugativem Transfer, und stabiler Vererbung involviert sind, reguliert. KorA bindet an sieben Operatorsequenzen (O_A1 - O_A7), die über das gesamte RP4-Plasmidgenom verteilt sind. Die Zielsetzung dieser Doktorarbeit war es, eine möglichst hochaufgelöste Kristallstruktur von KorA im Komplex mit seiner Operatorbindungsstelle zu erhalten, um den Mechanismus der Genregulation genau zu beleuchten. Die Struktur des Transkriptionsrepressors KorA, gebunden an ein pseudo-symmetrisches 18-bp DNA-Duplex O_A* (5'-C^{Br}U^{Br}U GTT TAG CTA AAC A^{Br}UT-3'), welches die symmetrische Operatorsequenz aufweist und 5-Bromodeoxyuridinnukleoside beinhaltet, wurde mittels MAD-Phasierung bei einer Auflösung von 1.96-Å bestimmt. Dies ist die erste publizierte Kristallstruktur des KorA Proteins.

Im Kristall ist KorA ein symmetrisches Dimeres, welches die DNA mit einem *Helix-Turn-Helix*-Motiv kontaktiert. Jede Hälfte der symmetrischen Operator-DNA bindet eine Kopie des Proteins in der großen Furche. Die Operator-DNA ist leicht gebogen und weist die Form von B-DNA auf. Die Oligonukleotide befinden sich in zwei Ausrichtungen im Kristall und überlagern sich in der Struktur. Aufgrund dieser zwei verschiedenen Orientierungen ist jeder spezifische Kontakt zum herauf- und herunterführenden DNA-Duplex ausgebildet, wobei sich die Längen der Wasserstoffbrückenbindungen einander sehr ähnlich sind. Durch Mutagenesestudien konnte die spezifische DNA-Erkennung durch KorA auf zwei Seitenketten von Arg48 und Gln53 eingegrenzt werden. Beide Reste bilden in der Kristallstruktur Wasserstoffbrückenbindungen zu insgesamt vier Basen innerhalb jedes

Halb-Operators aus. Die im Anschluss durchgeführten Gel-Retardationsexperimente konnten bestätigen, dass der im Kristall erhaltene Bindungszustand mit dem in Lösung gefundenen übereinstimmt. Gln53 wurde als der entscheidende Rest für die spezifische Operatorbindung ermittelt.

Ultrazentrifugationsexperimente bestätigten, dass KorA in Lösung ein sehr stabiles Dimeres ist. Weitere Untersuchungen betreffs des Bindungszustands ergaben, dass KorA DNA im molaren Verhältnis 2:1 bindet. Das hohe Bindungsvermögen von wt KorA wurde anhand von ITC-Bindungsstudien mit dem unbromierten 18-bp Oligonucleotid O_A (5'-CTT GTT TAG CTA AAC ATT-3'), welches die Operatorsequenz der Klasse I trägt, bestätigt. Kombiniert man die Daten aus Kristallstrukturanalyse und biochemischen Bindungsstudien, dann verdeutlicht sich, wie KorA die zwei Operatorklassen mit unterschiedlicher Affinität erkennt und bindet.

KorA besitzt ein einzigartiges Dimerisierungsmodul, welches auch bei zwei anderen RP4-Proteinen, TrbA und KlcB, zu finden ist. Berücksichtigt man die hohe Sequenzähnlichkeit ihrer C-Termini, kann angenommen werden, dass alle drei Proteine eine ähnliche Struktur der Dimerisierungsdomaine aufweisen. Kooperativität der globalen Repressorproteine KorA und TrbA mit KorB unterstützt diese Hypthese. Verdeutlicht man die konservierten Regionen auf der äußeren Molekülоberfläche von KorA, dann ergibt sich eine ausgedehnte Fläche identischer und ähnlicher Aminosäuren direkt auf dem oberen Teil der Dimerisierungsregion. Diese konservierte Region auf der Dimerisierungsoberfläche lässt die Schlussfolgerung zu, dass KorA, TrbA und KlcB mit dem globalen RP4-Repressorprotein KorB in gleicher Weise über eben dieses Dimerisierungsmodul kooperieren und dadurch die Stabilität des RP4 Plasmids regulieren.

Appendix A: References

Adamczyk, M. and Jagura-Burdzy, G. (2003) Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim Pol*, **50**, 425-453.

Austin, S. and Nordström, K. (1990) Partition-mediated incompatibility of bacterial plasmids. *Cell*, **60**, 351-354.

Bahl, M. I., Hansen, L. H., Goesmann, A. and Sorensen, S. J. (2007) The multiple antibiotic resistance IncP-1 plasmid pKJK5 isolated from a soil environment is phylogenetically divergent from members of the previously established alpha, beta and delta sub-groups. *Plasmid*, **58**, 31-43.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, **98**, 10037-10041.

Balzer, D., Ziegelin, G., Pansegrau, W., Kruft, V. and Lanka, E. (1992) KorB protein of promiscuous plasmid RP4 recognizes inverted sequence repetitions in regions essential for conjugative plasmid transfer. *Nucleic Acids Res*, **20**, 1851-1858.

Barnard, A., Wolfe, A. and Busby, S. (2004) Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr Opin Microbiol*, **7**, 102-108.

Batt, S. M., Bingle, L. E., Dafforn, T. R. and Thomas, C. M. (2009) Bacterial genome partitioning: N-terminal domain of IncC protein encoded by broad-host-range plasmid RK2 modulates oligomerisation and DNA binding. *J Mol Biol*, **385**, 1361-1374.

Behlke, J., Ristau, O. and Schonfeld, H. J. (1997) Nucleotide-dependent complex formation between the Escherichia coli chaperonins GroEL and GroES studied under equilibrium conditions. *Biochemistry*, **36**, 5149-5156.

Bhattacharyya, A. and Figurski, D. H. (2001) A small protein-protein interaction domain common to KlcB and global regulators KorA and TrbA of promiscuous IncP plasmids. *J Mol Biol*, **310**, 51-67.

Bingle, L. E., Macartney, D. P., Fantozzi, A., Manzoor, S. E. and Thomas, C. M. (2005) Flexibility in repression and cooperativity by KorB of broad host range IncP-1 plasmid RK2. *J Mol Biol*, **349**, 302-316.

Bingle, L. E., Rajasekar, K. V., Tul Muntaha, S., Nadella, V., Hyde, E. I. and Thomas, C. M. (2008) A single aromatic residue in transcriptional repressor protein KorA is critical for cooperativity with its coregulator KorB. *Mol Microbiol*, **70**, 1502-1514.

Bingle, L. E. and Thomas, C. M. (2001) Regulatory circuits for plasmid survival. *Curr Opin Microbiol*, **4**, 194-200.

Brennan, R. G. and Matthews, B. W. (1989) The helix-turn-helix DNA binding motif. *J Biol Chem*, **264**, 1903-1906.

Brenowitz, M., Pickar, A. and Jamison, E. (1991) Stability of a Lac repressor mediated "looped complex". *Biochemistry*, **30**, 5986-5998.

Browning, D. F. and Busby, S. J. (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol*, **2**, 57-65.

Burgess, R. R., Travers, A. A., Dunn, J. J. and Bautz, E. K. (1969) Factor stimulating transcription by RNA polymerase. *Nature*, **221**, 43-46.

Caspi, R., Pacek, M., Consiglieri, G., Helinski, D. R., Toukdarian, A. and Konieczny, I. (2001) A broad host range replicon with different requirements for replication initiation in three bacterial species. *EMBO J*, **20**, 3262-3271.

Chiu, C. M., Manzoor, S. E., Batt, S. M., Muntaha, S., Bingle, L. E. and Thomas, C. M. (2008) Distribution of the partitioning protein KorB on the genome of IncP-1 plasmid RK2. *Plasmid*, **59**, 163-175.

Clarage, J. B. and Phillips, G. N., Jr. (1994) Cross-validation tests of time-averaged molecular dynamics refinements for determination of protein structures by X-ray crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 24-36.

Collaborative Computational Project, No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-763.

Collado-Vides, J., Magasanik, B. and Gralla, J. D. (1991) Control site location and transcriptional regulation in Escherichia coli. *Microbiol Rev*, **55**, 371-394.

Correll, C. C., Freeborn, B., Moore, P. B. and Steitz, T. A. (1997) Use of chemically modified nucleotides to determine a 62-nucleotide RNA crystal structure: a survey of phosphorothioates, Br, Pt and Hg. *J Biomol Struct Dyn*, **15**, 165-172.

Courey, A. J. (2001) Cooperativity in transcriptional control. *Curr Biol*, **11**, R250-252.

Cowtan, K. (1994) 'dm': An automated procedure for phase improvement by density modification. *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr*, **31**, 34-38.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. and McPherson, A. (1994) Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr D Biol Crystallogr*, **50**, 414-423.

Datta, N. and Hedges, R. W. (1971) Compatibility groups among fi - R factors. *Nature*, **234**, 222-223.

deHaseth, P. L., Zupancic, M. L. and Record, M. T., Jr. (1998) RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol*, **180**, 3019-3025.

del Solar, G., Alonso, J. C., Espinosa, M. and Diaz-Orejas, R. (1996) Broad-host-range plasmid replication: an open question. *Mol Microbiol*, **21**, 661-666.

del Solar, G. and Espinosa, M. (2000) Plasmid copy number control: an ever-growing story. *Mol Microbiol*, **37**, 492-500.

DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific LLC, San Carlos, CA, USA.

Delbrück, H., Ziegelin, G., Lanka, E. and Heinemann, U. (2002) An Src homology 3-like domain is responsible for dimerization of the repressor protein KorB encoded by the promiscuous IncP plasmid RP4. *J Biol Chem*, **277**, 4191-4198.

Dennis, J. J. (2005) The evolution of IncP catabolic plasmids. *Curr Opin Biotechnol*, **16**, 291-298.

Diederichs, K. and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol*, **4**, 269-275.

Dmitrienko, V. E., Ishida, K., Kirfel, A. and Ovchinnikova, E. N. (2005) Polarization anisotropy of X-ray atomic factors and 'forbidden' resonant reflections. *Acta Crystallogr A*, **61**, 481-493.

Easter, C. L., Schwab, H. and Helinski, D. R. (1998) Role of the parCBA operon of the broad-host-range plasmid RK2 in stable plasmid maintenance. *J Bacteriol*, **180**, 6023-6030.

Ebersbach, G. and Gerdes, K. (2005) Plasmid segregation mechanisms. *Annu Rev Genet*, **39**, 453-479.

Ebright, R. H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol*, **304**, 687-698.

Evans, G. and Pettifer, R. F. (2001) CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J Appl Crystallogr*, **34**, 82-86.

Figurski, D. H., Young, C., Schreiner, H. C., Pohlman, R. F., Bechhofer, D. H., Prince, A. S. and D'Amico, T. F. (1985) Genetic interactions of broad host-range plasmid RK2: evidence for a complex replication regulon. *Basic Life Sci*, **30**, 227-241.

Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K. and Ramos, J. L. (1997) Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev*, **61**, 393-410.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) Protein indentification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, 571-607.

Gerdes, K., Moller-Jensen, J. and Bugge Jensen, R. (2000) Plasmid and chromosome partitioning: surprises from phylogeny. *Mol Microbiol*, **37**, 455-466.

Ghosh, S. K., Hajra, S., Paek, A. and Jayaram, M. (2006) Mechanisms for chromosome and plasmid segregation. *Annu Rev Biochem*, **75**, 211-241.

Gottesman, S. (1984) Bacterial regulation: global regulatory networks. *Annu Rev Genet*, **18**, 415-441.

Gralla, J. D. (1989) Bacterial gene regulation from distant DNA sites. *Cell*, **57**, 193-195.

Gralla, J. D. (1996) Activation and repression of E. coli promoters. *Curr Opin Genet Dev*, **6**, 526-530.

Gros, P., van Gunsteren, W. F. and Hol, W. G. (1990) Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. *Science*, **249**, 1149-1152.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*, **166**, 557-580.

Harrison, S. C. (1991) A structural taxonomy of DNA-binding domains. *Nature*, **353**, 715-719.

Hayes, F. and Barilla, D. (2006) The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nat Rev Microbiol*, **4**, 133-143.

Heinemann, J. A. and Sprague, G. F., Jr. (1989) Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature*, **340**, 205-209.

Heinemann, U., Büssow, K., Mueller, U. and Umbach, P. (2003) Facilities and methods for the high-throughput crystal structural analysis of human proteins. *Acc Chem Res*, **36**, 157-163.

Helmann, J. D. and Chamberlin, M. J. (1988) Structure and function of bacterial sigma factors. *Annu Rev Biochem*, **57**, 839-872.

Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51-58.

Hendrickson, W. A., Horton, J. R. and LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J*, **9**, 1665-1672.

Hendrickson, W. A., Smith, J. L., Phizackerley, R. P. and Merritt, E. A. (1988) Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. *Proteins*, **4**, 77-88.

Hooft, R. W., Vriend, G., Sander, C. and Abola, E. E. (1996) Errors in protein structures. *Nature*, **381**, 272.

Huffman, J. L. and Brennan, R. G. (2002) Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr Opin Struct Biol*, **12**, 98-106.

Jagura-Burdzy, G., Khanim, F., Smith, C. A. and Thomas, C. M. (1992) Crosstalk between plasmid vegetative replication and conjugative transfer: repression of the trfA operon by trbA of broad host range plasmid RK2. *Nucleic Acids Res*, **20**, 3939-3944.

Jagura-Burdzy, G., Kostelidou, K., Pole, J., Khare, D., Jones, A., Williams, D. R. and Thomas, C. M. (1999) IncC of broad-host-range plasmid RK2 modulates KorB transcriptional repressor activity In vivo and operator binding in vitro. *J Bacteriol*, **181**, 2807-2815.

Jagura-Burdzy, G. and Thomas, C. M. (1995) Purification of KorA protein from broad host range plasmid RK2: definition of a hierarchy of KorA operators. *J Mol Biol*, **253**, 39-50.

Jancarik, J., Scott, W. G., Milligan, D. L., Koshland, D. E., Jr. and Kim, S. H. (1991) Crystallization and preliminary X-ray diffraction study of the ligand-binding domain of the bacterial chemotaxis-mediating aspartate receptor of Salmonella typhimurium. *J Mol Biol*, **221**, 31-34.

Jones, S. and Thornton, J. M. (1997) Analysis of protein-protein interaction sites using surface patches. *J Mol Biol*, **272**, 121-132.

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, **47**, 110-119.

Kabsch, W. (1978) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr A*, **32**, 922-923.

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr*, **26**, 795-800.

Khare, D., Ziegelin, G., Lanka, E. and Heinemann, U. (2004) Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. *Nat Struct Mol Biol*, **11**, 656-663.

Kibbe, W. A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res*, **35**, W43-46.

Kini, R. M. and Evans, H. J. (1991) Molecular modeling of proteins: a strategy for energy minimization by molecular mechanics in the AMBER force field. *J Biomol Struct Dyn*, **9**, 475-488.

Kornacki, J. A., Chang, C. H. and Figurski, D. H. (1993) kil-kor regulon of promiscuous plasmid RK2: structure, products, and regulation of two operons that constitute the kilE locus. *J Bacteriol*, **175**, 5078-5090.

Kostelidou, K., Jagura-Burdzy, G. and Thomas, C. M. (1998) Mutational analysis of the global regulator KorA of broad-host-range plasmid RK2. *J Mol Biol*, **281**, 453-463.

Kostelidou, K., Jones, A. C. and Thomas, C. M. (1999) Conserved C-terminal region of global repressor KorA of broad-host-range plasmid RK2 is required for co-operativity between KorA and a second RK2 global regulator, KorB. *J Mol Biol*, **289**, 211-221.

Kostelidou, K. and Thomas, C. M. (2002) DNA recognition by the KorA proteins of IncP-1 plasmids RK2 and R751. *Biochim Biophys Acta*, **1576**, 110-118.

Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, **372**, 774-797.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Lamzin, V. S., Perrakis, A. and Wilson, K. S. (2001) The ARP/wARP suite for automated construction and refinement of protein models. *International Tables for Crystallography. Volume F: Crystallography of Biological Macromolecules (Rossmann, M. G. and Arnold, E. eds.), Dordrecht, Kluwer Academic Publishers, The Netherlands*, 720-722.

Lanzer, M. and Bujard, H. (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A*, **85**, 8973-8977.

Larsen, M. H. and Figurski, D. H. (1994) Structure, expression, and regulation of the kilC operon of promiscuous IncP alpha plasmids. *J Bacteriol*, **176**, 5022-5032.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*, **26**, 283-291.

Lavery, R. and Sklenar, H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn*, **6**, 63-91.

Lederberg, J. (1952) Cell genetics and hereditary symbiosis. *Physiol Rev*, **32**, 403-430.

Lederberg, J. (1998) Plasmid (1952-1997). *Plasmid*, **39**, 1-9.

Lloyd, G., Landini, P. and Busby, S. (2001) Activation and repression of transcription initiation in bacteria. *Essays Biochem*, **37**, 17-31.

Lowbury, E. J., Lilly, H. A., Kidson, A., Ayliffe, G. A. and Jones, R. J. (1969) Sensitivity of Pseudomonas aeruginosa to antibiotics: emergence of strains highly resistant to carbenicillin. *Lancet*, **2**, 448-452.

McPherson, A. (1999) Crystallization of Biological Macromolecules. Cold Spring Harbor Laboratory Press: Cold Spring Harbor.

Merika, M. and Thanos, D. (2001) Enhanceosomes. *Curr Opin Genet Dev*, **11**, 205-208.

Meyer, R. and Hinds, M. (1982) Multiple mechanisms for expression of incompatibility by broad-host-range plasmid RK2. *J Bacteriol*, **152**, 1078-1090.

Mullis, K. B. and Faloona, F. A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*, **155**, 335-350.

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. and Darst, S. A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*, **296**, 1285-1290.

Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.

Nesterenko, M. V., Tilley, M. and Upton, S. J. (1994) A simple modification of Blum's silver stain method allows for 30 minute detection of proteins in polyacrylamide gels. *J Biochem Biophys Methods*, **28**, 239-242.

Nordström, K., Molin, S. and Light, J. (1984) Control of replication of bacterial plasmids: genetics, molecular biology, and physiology of the plasmid R1 system. *Plasmid*, **12**, 71-90.

Novick, R. P. (1987) Plasmid incompatibility. *Microbiol Rev*, **51**, 381-395.

Olsen, R. H. and Shipley, P. (1973) Host range and properties of the Pseudomonas aeruginosa R factor R1822. *J Bacteriol*, **113**, 772-780.

Pabo, C. O. and Sauer, R. T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem*, **61**, 1053-1095.

Pansegrau, W. and Lanka, E. (1987) Conservation of a common 'backbone' in the genetic organization of the IncP plasmids RP4 and R751. *Nucleic Acids Res*, **15**, 2385.

Pansegrau, W., Lanka, E., Barth, P. T., Figurski, D. H., Guiney, D. G., Haas, D., Helinski, D. R., Schwab, H., Stanisich, V. A.*, et al.* (1994) Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J Mol Biol*, **239**, 623-663.

Pape, T. and Schneider, T. R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs. *J Appl Crystallogr*, **37**, 843-844.

Pérez-Rueda, E. and Collado-Vides, J. (2000) The repertoire of DNA-binding transcript-tional regulators in Escherichia coli K-12. *Nucleic Acids Res*, **28**, 1838-1847.

Pérez-Rueda, E., Gralla, J. D. and Collado-Vides, J. (1998) Genomic position analyses and the transcription machinery. *J Mol Biol*, **275**, 165-170.

Roberts, R. C., Strom, A. R. and Helinski, D. R. (1994) The parDE operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss. *J Mol Biol*, **237**, 35-51.

Rojo, F. (1999) Repression of transcription initiation in bacteria. *J Bacteriol*, **181**, 2987-2991.

Rosche, T. M., Siddique, A., Larsen, M. H. and Figurski, D. H. (2000) Incompatibility protein IncC and global regulator KorB interact in active partition of promiscuous plasmid RK2. *J Bacteriol*, **182**, 6014-6026.

Ross, W., Ernst, A. and Gourse, R. L. (2001) Fine structure of E. coli RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev*, **15**, 491-506.

Sanderson, A., Mitchell, J. E., Minchin, S. D. and Busby, S. J. (2003) Substitutions in the Escherichia coli RNA polymerase sigma70 factor that affect recognition of extended -10 elements at promoters. *FEBS Lett*, **544**, 199-205.

Sanishvili, R., Besnard, C., Camus, F., Fleurant, M., Pattison, P., Bricogne, G. and Schiltz, M. (2007) Polarization-dependence of anomalous scattering in brominated DNA and RNA molecules, and importance of crystal orientation in single- and multiple-wavelength anomalous diffraction phasing. *J Appl Crystallogr*, **40**, 552-558.

Schiltz, M. and Bricogne, G. (2008) Exploiting the anisotropy of anomalous scattering boosts the phasing power of SAD and MAD experiments. *Acta Crystallogr D Biol Crystallogr*, **D64**, 711-729.

Schiltz, M., Dumas, P., Ennifar, E., Flensburg, C., Paciorek, W., Vonrhein, C. and Bricogne, G. (2004) Phasing in the presence of severe site-specific radiation damage through dose-dependent modelling of heavy atoms. *Acta Crystallogr D Biol Crystallogr*, **60**, 1024-1031.

Schmidhauser, T. J. and Helinski, D. R. (1985) Regions of broad-host-range plasmid RK2 involved in replication and stable maintenance in nine species of gram-negative bacteria. *J Bacteriol*, **164**, 446-455.

Schneider, T. R. and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr*, **58**, 1772-1779.

Schreiner, H. C., Bechhofer, D. H., Pohlman, R. F., Young, C., Borden, P. A. and Figurski, D. H. (1985) Replication control in promiscuous plasmid RK2: kil and kor functions affect expression of the essential replication gene trfA. *J Bacteriol*, **163**, 228-237.

Seeman, N. C., Rosenberg, J. M. and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc Natl Acad Sci U S A*, **73**, 804-808.

Senear, D. F., Ross, J. B. and Laue, T. M. (1998) Analysis of protein and DNA-mediated contributions to cooperative assembly of protein-DNA complexes. *Methods*, **16**, 3-20.

Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr A*, **64**, 112-122.

Siddique, A. and Figurski, D. H. (2002) The active partition gene incC of IncP plasmids is required for stable maintenance in a broad range of hosts. *J Bacteriol*, **184**, 1788-1793.

Sorenson, M. K., Ray, S. S. and Darst, S. A. (2004) Crystal structure of the flagellar sigma/anti-sigma complex sigma(28)/FlgM reveals an intact sigma factor in an inactive conformation. *Mol Cell*, **14**, 127-138.

Spassky, A., Kirkegaard, K. and Buc, H. (1985) Changes in the DNA structure of the lac UV5 promoter during formation of an open complex with Escherichia coli RNA polymerase. *Biochemistry*, **24**, 2723-2731.

Summers, D. K., Beton, C. W. and Withers, H. L. (1993) Multicopy plasmid instability: the dimer catastrophe hypothesis. *Mol Microbiol*, **8**, 1031-1038.

Summers, D. K. and Sherratt, D. J. (1984) Multimerization of high copy number plasmids causes instability: CoIE1 encodes a determinant essential for plasmid monomerization and stability. *Cell*, **36**, 1097-1103.

Suzuki, M. and Brenner, S. E. (1995) Classification of multi-helical DNA-binding domains and application to predict the DBD structures of sigma factor, LysR, OmpR/PhoB, CENP-B, Rapl, and Xy1S/Ada/AraC. *FEBS Lett*, **372**, 215-221.

Taylor, D. E. and Bradley, D. E. (1987) Location on RP4 of a tellurite resistance determinant not normally expressed in IncP alpha plasmids. *Antimicrob Agents Chemother*, **31**, 823-825.

Templeton, L. K. and Templeton, D. H. (1995) Polarized dispersion for X-rays scattered by an aromatic bromide. *J Synchrotron Radiat*, **2**, 31-35.

Theophilus, B. D., Cross, M. A., Smith, C. A. and Thomas, C. M. (1985) Regulation of the trfA and trfB promoters of broad host range plasmid RK2: identification of sequences essential for regulation by trfB/korA/korD. *Nucleic Acids Res*, **13**, 8129-8142.

Thomas, C. M., Ibbotson, J. P., Wang, N. Y., Smith, C. A., Tipping, R. and Loader, N. M. (1988) Gene regulation on broad host range plasmid RK2: identification of three novel operons whose transcription is repressed by both KorA and KorC. *Nucleic Acids Res*, **16**, 5345-5359.

Thomas, C. M. and Smith, C. A. (1986) The trfB region of broad host range plasmid RK2: the nucleotide sequence reveals incC and key regulatory gene trfB/korA/korD as overlapping genes. *Nucleic Acids Res*, **14**, 4453-4469.

Thomas, C. M., Smith, C. A., Shingler, V., Cross, M. A., Hussain, A. A. and Pinkney, M. (1985) Regulation of replication and maintenance functions of broad host-range plasmid RK2. *Basic Life Sci*, **30**, 261-276.

Thomas, C. M., Theophilus, B. D., Johnston, L., Jagura-Burdzy, G., Schilf, W., Lurz, R. and Lanka, E. (1990) Identification of a seventh operon on plasmid RK2 regulated by the korA gene product. *Gene*, **89**, 29-35.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.

Thorsted, P. B., Macartney, D. P., Akhtar, P., Haines, A. S., Ali, N., Davidson, P., Stafford, T., Pocklington, M. J., Pansegrau, W.*, et al.* (1998) Complete sequence of the IncPbeta plasmid R751: implications for evolution and organisation of the IncP backbone. *J Mol Biol*, **282**, 969-990.

Tran, J. H. and Jacoby, G. A. (2002) Mechanism of plasmid-mediated quinolone resistance. *Proc Natl Acad Sci U S A*, **99**, 5638-5642.

Uhlin, B. E. and Nordström, K. (1978) A runaway-replication mutant of plasmid R1drd-19: temperature-dependent loss of copy number control. *Mol Gen Genet*, **165**, 167-179.

Valentine, S. A., Chen, G., Shandala, T., Fernandez, J., Mische, S., Saint, R. and Courey, A. J. (1998) Dorsal-mediated repression requires the formation of a multiprotein repression complex at the ventral silencer. *Mol Cell Biol*, **18**, 6584-6594.

Vannini, A., Volpari, C., Gargioli, C., Muraglia, E., Cortese, R., De Francesco, R., Neddermann, P. and Marco, S. D. (2002) The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J*, **21**, 4393-4401.

Wagner, E. G. and Brantl, S. (1998) Kissing and RNA stability in antisense control of plasmid replication. *Trends Biochem Sci*, **23**, 451-454.

Wang, W. and Malcolm, B. A. (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *Biotechniques*, **26**, 680-682.

Waters, V. L. (2001) Conjugation between bacterial and mammalian cells. *Nat Genet*, **29**, 375-376.

Williams, D. R., Macartney, D. P. and Thomas, C. M. (1998) The partitioning activity of the RK2 central control region requires only incC, korB and KorB-binding site O(B)3 but other KorB-binding sites form destabilizing complexes in the absence of O(B)3. *Microbiology*, **144**, 3369-3378.

Williams, D. R., Motallebi-Veshareh, M. and Thomas, C. M. (1993) Multifunctional repressor KorB can block transcription by preventing isomerization of RNA polymerase-promoter complexes. *Nucleic Acids Res*, **21**, 1141-1148.

Wilson, J. W., Sia, E. A. and Figurski, D. H. (1997) The kilE locus of promiscuous IncP alpha plasmid RK2 is required for stable maintenance in Pseudomonas aeruginosa. *J Bacteriol*, **179**, 2339-2347.

Winn, M. D., Isupov, M. N. and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr*, **57**, 122-133.

Wintjens, R. and Rooman, M. (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J Mol Biol*, **262**, 294-313.

Wiseman, T., Williston, S., Brandts, J. F. and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*, **179**, 131-137.

Wösten, M. M. (1998) Eubacterial sigma-factors. *FEMS Microbiol Rev*, **22**, 127-150.

Young, C., Prince, A. S. and Figurski, D. H. (1985) korA function of promiscuous plasmid RK2: an autorepressor that inhibits expression of host-lethal gene kilA and replication gene trfA. *Proc Natl Acad Sci U S A*, **82**, 7374-7378.

Appendix B: Directories

Figure	Title	Page
1.1	RNA polymerase bound to a promoter site.	2
1.2	Mechanisms for transcription activation.	3
1.3	Mechanisms for transcription repression.	4
1.4	Cooperativity between transcription regulators.	6
1.5	Map of plasmid RP4.	10
1.6	The *korAB* operon of RP4.	11
3.1	Testing for DNA binding of KorA.	36
3.2	ITC thermogram analysis.	40
3.3	Protein crystallization phase diagram.	42
3.4	Hanging drop vapor diffusion method.	43
3.5	Bragg's law.	45
3.6	The Ewald sphere.	46
3.7	The structure factor.	48
3.8	The Friedel pairs.	49
3.9	Anomalous scattering.	50
3.10	The Harker construction.	51
4.1	SDS PAGE showing the first purification steps of KorA.	55
4.2	Affinity chromatography of partially pure KorA.	56
4.3	Cation-exchange chromatography of KorA.	56
4.4	Gel filtration elution profile of KorA.	57
4.5	SDS PAGE showing the final purification steps of KorA.	57
4.6	Mass determination of KorA in solution using analytical ultracentrifugation.	58
4.7	Gel filtration elution profile of the KorA-O_A* complex.	59
4.8	PAGE evaluation of the gel filtration run of KorA-O_A* complex.	59
4.9	Crystals of the KorA-O_A* complex.	60
4.10	Anomalous scattering curves.	61
4.11	Crystal packing of the KorA-O_A* complex.	64
4.12	The KorA monomer.	65
4.13	The KorA dimer.	66
4.14	Interactions within the dimerization module.	67
4.15	Stabilization of the dimerization module.	68

Figure	Title	Page
4.16	The DNA-binding domain.	68
4.17	Orientation of the operator binding site.	69
4.18	The operator binding site.	70
5.1	Polarization-dependence of anomalous scattering.	72
5.2	Crystal packing of the brominated O_A* oligonucleotides.	73
5.3	Specific binding of KorA to its operator binding site.	76
5.4	Structural comparison.	77
5.5	KorA homologs from IncP-1 plasmids.	78
5.6	The KorA dimerization module.	81
5.7	Possible interaction between KorA and KorB at RP4 promoters.	84

Table	Title	Page
1.1	Plasmid incompatibility groups in *P. aeruginosa*.	8
1.2	The seven O_A sites on RP4.	12
2.1	Hampton Crystal Screen 1 formulations.	23
2.2	Hampton Crystal Screen 2 formulations.	24
3.1	Standard PCR setup.	27
3.2	Standard PCR program.	28
3.3	Molar extinction coefficients derived from *OligoCalc*.	29
3.4	Standard QCM reaction setup.	30
3.5	Standard QCM program.	30
3.6	Molar extinction coefficients calculated with *ProtParam*.	35
3.7	Silver staining protocol.	37
4.1	Anomalous scattering factors at peak and inflection wavelength.	61
4.2	Data collection statistics.	62
4.3	Refinement statistics.	64
5.1	Specific binding of Arg48 and Gln53 to O_A.	76
5.2	Conservation of KorA operators between RP4 and R751.	79

Appendix C: Abbreviations

°	degree
°C	degree Celsius
Å	Ångström
λ	wavelength
ε	molar extinction coefficient
µ	micro
A	absorbance, Ampère, adenine
aa	amino acid
AA	acrylamide
Amp	ampicillin
ASP	ammonium sulfate precipitation
APBS	adaptive Poisson-Boltzmann solver
APS	ammonium persulfate
a.u.	absorbance units
BESSY	Berlin Electron Storage Ring Company for Synchrotron Radiation
Bor	boron
bp	base pair(s)
^{Br}U	5-bromo-deoxyuridine
c	centi
C	cysteine, concentration
Cb	carbenicillin
Cm	chloramphenicol
Cma	chromosome mobilizing ability
CTAB	hexadecyltrimethylammonium bromide
CTD	C-terminal domain
Da	Dalton
ddH_2O	destilled, millipore water
DM	density modification
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
DTT	1,4-dithiothreitol
E.	*Escherichia*
EDTA	ethylene diamine tetraacetate
EMSA	electrophoretic mobility shift assay
et al.	*et alii* (and others)

FPLC	fast performance liquid chromatography
g	gram
g	g-force
G	guanine
GC	gel filtration chromatography
Gm	gentamicin
h	hour
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid
Hg	mercury
HPLC	high performance liquid chromatography
HTH	helix-turn-helix
Inc	incompatibility
IPTG	isopropyl-1-thio-β-D-galactopyranoside
ITC	isothermal titration calorimetry
k	kilo
K	Kelvin
k_{+1}	association rate constant
k_{-1}	dissociation rate constant
K_A	equilibrium association constant
kb	kilobases
K_D	equilibrium dissociation constant
Km	kanamycin
l	liter, length
LB	Luria-Bertani
m	milli or meter
M	mega or molar
MAD	Multi-wavelength anomalous dispersion
min	minute
MME	monomethyl ether
MPD	2-methyl-2,4-pentanediol
n	nano
N	stoichiometry
NCS	non-crystallographic symmetry
NTD	N-terminal domain
O_A	operator sequence of KorA
O_B	operator sequence of KorB
OD	optical density
P.	*Pseudomonas*

Pa	Pascal
PAGE	polyacrylamide gel electrophoresis
par	partition
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PEI	polyethyleneimine
pH	potentia hydrogenii, measure for acidity in aqueous solutions
Phi	phage inhibition
pI	isolelectric point
Pmr	phenyl mercuric acetate
psi	pound per square inch
QCM	QuikChange mutagenesis
rmsd	Root-mean-square deviation
RNA	ribonucleic acid
RNAP	RNA polymerase
rpm	rotations per minute
s or sec	second
SAD	Single-wavelength anomalous dispersion
SDS	sodium dodecylsulfate
Sm	streptomycin
Su	sulfonamides
T	thymine, temperature
TBE	Tris-borate-EDTA
Tc	tetracycline
TCA	trichloro acetate
Te	potassium tellurite
TEMED	N,N,N’,N’-tetramethylethylenediamine
TLS	translation, libration, screw-motion
Tm	trimethoprim
TRIS	Tris(hydroxymethyl)-aminomethan
U	unit
UV	ultraviolet
v/v	volume per volume
v/w	volume per weight
V	Volt
wt	wild-type

Curriculum vitae

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.

Danksagung

Ein großes Dankeschön an alle, die mich während meiner Doktorarbeit auf vielerlei Art unterstützt haben. Ich bin sehr dankbar für dieses Thema, den fachlichen Beistand und guten Rat, die vielen Tipps und Tricks, diese schöne Zeit und natürlich dafür, dass es Euch gibt.

Udo Heinemann, Jürgen Joachim Müller, Ulrich Gohlke, Klaas Max, Kerstin Böhm, Daniel Kümmel, Claudia Maria Haas, Sarbani Bhattacharya, Harald Striegl, Cheng-Cheng Wang, Jennifer Hanna, Dheeraj Khare, Anette Feske, Birgit Cloos, Andreas Knespel, Janine Mikutta, Maria Buchsteiner, Katja Scholz, Anna Szecsenyi-Nagy, Friederike Skole, Mandy Terne

Erich Lanka, Joachim Behlke, Yvette Roske, Anja Schütz, Ulf Lenski, Mathias Höschen, Uwe Müller, Jörg Schulze, Eva-Christina Müller, Albrecht Otto, Sandro Keller, Natalie Bordag, Nadin Jahnke, Gerdi Hölzl, Anja Sieber, Monika Georgi

De Muttis

Meine liebe Familie

Jörg Schiller