Crystallographic and Functional Study on DNA Binding Proteins: Repressor and Partitioning Protein KorB from RP4 Plasmid and the Transposase 'Sleeping Beauty' of Vertebrate Origin

Inaugural Dissertation
zur Erlangung des akademischen Grades

´´Doktor der Naturwissenschaften´´
- Dr. rer. nat. -

am Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

Vorgelegt von M.Sc., MPhil (Res)
Dheeraj Khare
aus Lucknow, Indien

Berlin 2004

Diese Arbeit wurde in Zeitraum von Oktober 1999 bis Februar 2004 in der Forschungsgruppe Kristallographie (Leiter: Prof. Dr. Udo Heinemann) am Max-Delbrück-Centrum für Molekulare Medizin in Berlin angefertigt.

1. Gutachter: Prof. Dr. Udo Heinemann
2. Gutachter: Prof. Dr. Volker A. Erdmann

Eingereicht am: 23. Februar 2004
Tag der mündlichen Prüfung am: 15. Juni 2004

Table of Contents

Abstract

DNA-binding proteins play a central role in all aspects of genetic activity within an organism, such as transcription, DNA rearrangement, replication and repair. It is therefore extremely important to analyse the nature of the complexes that are formed between proteins and DNA, as they form the basis of our knowledge of how these processes occur.

Partitioning of low copy-number plasmids at cell division is quite similar to mitosis in that, before cell division, paired plasmid molecules are separated from each other and actively moved apart. There are three components that are essential for partitioning to occur, two plasmid encoded proteins and a *cis*-acting centromere-like site on the plasmid DNA. The first gene of the partition operon involves either an ATPase with Walker-type ATP binding motifs (ParA) or an evolutionary unrelated actin-type-ATPase (ParM). The second partition protein (ParB or ParR) binds to the cis-acting DNA partition site and recruits the ATPase into the nucleoprotein complex. Genes for homologues of ParA and ParB exist on many bacterial chromosomes and plasmids.

The IncPα plasmids, exemplified by the low copy-number plasmid RP4, encode for an active partitioning system with homologues of ParA and ParB partitioning proteins called IncC, a putative ATPase, and KorB, a specific DNA-binding protein, respectively. Although KorB has a net negative charge (pI = 4.6), it recognizes and binds specifically to palindromic operator O_B (consensus sequence 5′ TTTAGC$^{G}/_{C}$GCTAAA 3′) occurring at 12 different sites on the plasmid (operators O_B1 to O_B12), 6 of which are involved in transcriptional regulation. The KorB protein acts as a multifunctional regulator in control circuits of plasmid housekeeping genes involved in replication, maintenance and conjugation. Additionally, KorB functions as the ParB homologue of the plasmid's partitioning system.

The major part of the work presented in this thesis concerns the crystal structure of the DNA-binding domain of a global regulator, KorB. This work reveals for the first time the structural basis of DNA binding of a family member of the ParB DNA partitioning factors to its recognition sequence.

In this thesis, the X-ray crystal structure of the complex comprising the DNA-binding domain of KorB (KorB-O) and its operator sequence O_B, determined to 2.2-Å resolution, is presented. The KorB-O—O_B complex crystallized in spacegroup $P3_221$ with cell parameters *a* = 110.44 Å, *c*= 160.53 Å, with two copies of the complex in the asymmetric unit. Each half-site of the palindromic operator DNA binds one copy of the protein into the major groove. The KorB-O bound operator DNA adopts a standard B-DNA conformation with a straight helix axis. The protein structure consists of eight helices two of which belong to a canonical helix-turn-helix DNA-binding motif. The α3-turn-α4 segment of KorB-O has a sequence signature and conformation typical for a helix-turn-helix motif generally involved in specific DNA major groove interactions. However, residues from these two helices do not form direct base contacts in the O_B region.

The structure reveals mainly two residues, Thr211 and Arg240, which recognize the O_B sequence through direct hydrogen bonding. This was further confirmed by mutagenesis, where mutant KorB proteins Thr211Ala and Arg240Ala do not recognize O_B specifically suggesting that the mode of operator binding by the KorB-O fragment observed in the crystal structure reflects the specific DNA binding of the wildtype protein. The outer surface of the DNA-bound KorB-O mirrors the overall acidity of KorB, whereas DNA binding occurs *via* a basic interaction surface. A model of KorB including the structure of its dimerization domain is presented that considers the interaction with the highly basic ParA homologue IncC.

The last chapter of the thesis describes work done to characterize and crystallize the DNA-binding domain of the transposase '*Sleeping Beauty*' of vertebrate origin. Efforts were made to purify and crystallize different constructs of the N-terminal DNA binding domain. In particular, a protocol is presented for the purification of an active form of the full-length transposase.

Zusammenfassung

DNA-bindende Proteine spielen für alle Aspekte der genetischen Aktivität eines Organismus, wie bei der Transkription, der Rekombination der DNA, ihrer Replikation und ihrer Reparatur eine zentrale Rolle. Sie müssen die DNA sequenzspezifisch erkennen und mit ihr wechselwirken, so dass die in der DNA verschlüsselte Information abgelesen werden kann. Demzufolge ist es sehr wichtig, die Natur von DNA-Protein-Komplexen zu analysieren. Ihre Strukturen bilden die Voraussetzung für das Verständnis der genannten Prozesse.

Die Partitionierung von Plasmiden geringen Kopienzahl bei der Zellteilung ist dem Prozess der Mitose dahin gehend verwandt, dass gepaarte Plasmidmoleküle von einander getrennt und aktiv räumlich separiert werden. Drei Komponenten sind für die Teilung notwendig: zwei plasmidkodierte Proteine und ein *cis*-aktiver Zentromer-ähnlicher Abschnitt der Plasmid-DNA. Das erste Gen auf dem Partitions-Operon kodiertt entweder für eine ATPase mit Walker-ähnlichem ATP-Bindungsmotiv (ParA) oder eine evolutionär nicht verwandte Aktin-ähnliche ATPase (ParM). Das zweite Partitionsprotein (ParB oder ParR) bindet an den *cis*-aktiven DNA-Partitionierungsabschnitt und erlaubt der ATPase die Bindung an den Nukleo-Proteinkomplex. Gene für ParA- und ParB-Homologe existieren auf vielen bakteriellen Chromosomen und Plasmiden.

Die IncPα-Plasmide, hier am Beispiel des "low-copy-number"-Plasmids RP4 erläutert, kodieren für ein aktives Partitionierungssystem. Seine Homologe der Teilungsproteine ParA und ParB sind IncC, eine mutmaßliche ATPase und KorB, ein sequenzspezifisch DNA-bindendes Protein. Trotz seiner negativen Gesamtladung (pI = 4.6), erkennt und bindet KorB spezifisch die palindromische Operatorsequenz O_B (Konsensussequenz 5´ TTTAGC$^{G}/_{C}$GCTAAA 3´) an 12 verschiedenen Orten des Plasmids (Operatoren O_B1 bis O_B12). Sechs dieser Abschnitte sind in die Transkriptionsregulierung involviert. Die KorB-Proteine wirken als multifunktionelle Regulatoren der Kontrollabschnitte von Plasmidgenen, welche an der Replikation, Erhaltung und Konjugation beteiligt sind. KorB stellt das ParB-Homologe im Partitionierungssystem des Plasmids dar.

Der Hauptteil dieser Arbeit widmet sich der Kristallstruktur der DNA-bindenden Domäne des globalen Regulators KorB. Es wird erstmalig die strukturelle Grundlage der DNA-Bindung eines Proteins der ParB-Familie an seine Erkennungssequenz beschrieben. Die Röntgenkristallstruktur des Komplexes zwischen der DNA-Bindungsdomäne von KorB (KorB-O) und seiner Operatorsequenz O_B wurde bei 2.2 Å Auflösung entschlüsselt. Jede Hälfte der palindromischen Operator-DNA bindet eine Kopie des Proteins in der groβen Furche. Die Proteinstruktur umfasst acht Helices, von denen zwei zu einem "Helix-Turn-Helix"-Motiv (HTH) gehören. Das α3-turn-α4-Segment des KorB-O besitzt eine Sequenzsignatur und Konformation, die für Proteine typisch ist, welche spezifische Interaktionen innerhalb der großen Furche doppelsträngiger DNA eingehen. Aminosäureseitenketten des HTH-Motivs bilden jedoch keinen direkten Kontakt mit Basenpaaren der O_B-Region aus.

Es wurden vielmehr zwei Reste, Thr211 und Arg240, außerhalb des HTH-Motivs identifiziert, die die O_B Sequenz durch direkte Wasserstoffbrückenbindungen erkennen. Dies wurde durch gerichtete Mutagenese bestätigt: Die Proteinvarianten Thr211Ala und Arg240Ala sind nicht in der Lage, O_B spezifisch zu erkennen. Dies weist darauf hin, dass die Art der Operatorbindung durch das KorB-O Fragment, die in der Kristallstruktur beobachtet wurde, die spezifische DNA-Bindung des intakten Wildtyp-Proteins widerspiegelt. Die DNA-abgewandte Oberfläche des KorB-O trägt eine negative Nettoladung, wohingegen die DNA-Bindung über eine basische Oberfläche erfolgt. Ein die C-terminale Dimerisierungsdomäne einschließendes Modell des KorB-DNA-Komplexes zeigt einen vollständigen Einschluss des DNA-Doppelstrangs durch das Protein und erlaubt eine Diskussion seiner Wechselwirkung mit dem stark basischen ParA-Homologen IncC.

Das letzte Kapitel dieser Arbeit fasst Arbeitsschritte zur Charakterisierung und Kristallisation der DNA-Bindungsdomäne der Transposase "Sleeping Beauty" zusammen. Neben der Reinigung und Kristallisation verschiedener Konstrukte der N-terminalen DNA-Bindungsdomäne dieser Vertebraten-Transposase wurde vor allem ein Protokoll für die Reinigung einer aktiven Form der intakten Transposase etabliert.

Acknowledgements

There are a lot of people whom I would like to thank for a variety of reasons and without whom this thesis would not have been possible.

Firstly, I would like to thank my supervisor Prof. Dr. Udo Heinemann for being patient and supporting this work with knowledge, ideas and criticism. I owe much of my insight and interest in crystallography to the inspiring lectures and guidance given by Udo.

I was fortunate to have nice collaborators like Erich Lanka, Günter Ziegelin, Heinz Welfle, Zoltan Ivics, Lubomir Dostal, Hatem Zayed and Joachim Behlke. All three, Udo, Erich and Günter were a great help to reach the final version of the manuscript on KorB. Zsuzsa and Zoltan helped with many useful discussions.

I am grateful to all the staff in the Department, Jürgen for sorting all the crystallography related problems, Anette for helping in the lab, Carola for technical assistance, Andreas for all the technical help with the machines and not the least Birgit, who was always ready to help. Also I would like to appreciate the help of Yves Müller who is a wonderful person and a great teacher.

My sincere thanks are to my colleagues for always creating a pleasant working atmosphere in the lab and helping with problems.

I am also thankful to the many friends who have shared with me the many moments of joy and encouraged me during the occasional bouts of frustration I have experienced. Above all my stay in Berlin was a really happy one.

A special thanks goes to Zhyldyz for careful reading the thesis and her helpful and critical comments that improved the introduction a lot.

My greatest thanks are reserved for my parents, brother, sister in law and also my small niece Kavya. Their belief in my abilities and me has allowed me to achieve my goals. For this and their love I owe them a debt of thanks.

I would also like to thank the MDC, Graduate College and DFG for providing the resources and funding for my research throughout this PhD.

Thank you all!!

Dheeraj

Abbreviations

Å	Angstrom, 1 Å = 10^{-10} m
^{Br}U	deoxybromouridine
bp	base pair
DEAE	diethylaminoethyl Sephacel
DR	direct repeat
DTT	dithiothreitol
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
FPLC	fast performance liquid chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HTH	helix-turn-helix
IPTG	isopropyl-β-D-thiogalactopyranoside
IR	inverted repeat
Inc	incompatibility
KorB-O	DNA (operator) binding domain of KorB
MR	molecular replacement
MIR	multiple isomorphous replacement
MAD	multiwavelength anomalous diffraction
MPD	2-methyl-2,4-pentanediol
mAU	milli-absorption units
ml	millilitre
mM	millimolar
µl	microlitre
NCS	non-crystallographic symmetry
O_B	operator DNA
PAGE	polyacrylamide gel electrophoresis
Par	partitioning
PCR	polymerase chain reaction
PDB	protein data bank

pI	isoelectric point
rmsd	root mean square deviation
SB	*Sleeping Beauty*
SDS	sodium dodecyl sulfate
Tris	Trishydroxymethylaminomethane
UV	ultra violet
wt/vol	weight/volume