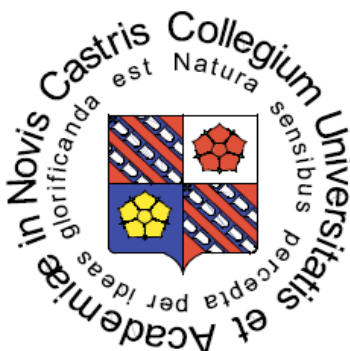


UNIVERSITY OF SOUTH BOHEMIA
INSTITUTE OF PHYSICAL BIOLOGY
and
INSTITUTE OF SYSTEMS BIOLOGY AND ECOLOGY
ACADEMY OF SCIENCES OF THE CZECH REPUBLIC



Mikalai Lapkouski

Structure of the motor subunit and translocation model for EcoR124I restriction-modification complex

Ph.D. THESIS

Supervisor:

Cséfalvay Eva, Ph. D.

Dept. of Structure and Function of Proteins,
Institute of Systems Biology and Ecology,
Academy of Sciences of the Czech Republic,
Institute of Physical Biology,
University of South Bohemia

Scientific advisor:

Assoc. Prof. **Rudiger Ettrich, Ph. D.**

Dept. of Structure and Function of Proteins,
Institute of Systems Biology and Ecology,
Academy of Sciences of the Czech Republic,
Institute of Physical Biology,
University of South Bohemia

Nové Hradky, 2008

PROHLÁŠENÍ

Prohlašuji, že svoji disertační práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v nezkrácené podobě, elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách.

V Nových Hradech, 16. září 2008

.....

Podpis studenta

ACKNOWLEDGEMENTS

I thank in the first case to all people who greatly contributed to the research work which has been done over the last three years in the fascinating field of DNA interacting proteins in the structural biology group at the **Dept. Structure and Function of Proteins**, a joint department of the Institute of Systems Biology and Ecology, Academy of Sciences, and the Institute of Physical Biology, University of South Bohemia. I greatly appreciate all valuable comments, offers, suggestions and of course hand on help. Particularly I am extremely grateful to the following outstanding scientists in their fields:

Eva Cséfalvay Ing. Ph.D. for noticing me during the Summer School 2005 and hosting as a student under her supervision for all this time since then. Close cooperation and mutually solving different problems helped me a lot in acquiring a practical experience and knowledge in the field of molecular biology.

Rudiger Ettrich Assoc. Prof., RNDr. Ph.D., head of the department, for helping and teaching in analyzing results and introducing me to the broad and powerful field of computational biology. A special thanks I want to say to him for bringing me into the top scientific community by taking me to the international meetings and conferences and introducing to the famous people in science.

Santosh Panjekar Ph.D., staff scientist at EMBL Hamburg Outstation for a long time assistance during the data collections and shearing his experience in determining atomic structures of macromolecules. Also, for hosting me as a visiting student at EMBL for many times.

Kutá-Smatanová Ivana, Ph.D., Mgr., the head of the Laboratory of Crystallogenesi and Biomolecular Crystallography in Nove Hradý for providing a practically convenient environment for working with macromolecular crystals and valuable consultations in the field.

Jannette Carey Prof. of Princeton University in New Jersey, for an enormous help in analyzing scientific results and tremendous help in theoretical chemistry and structural biology. I was shown how important is to thoroughly and rigorously analyze results and make them clearly interpretable for others

My profound gratitude also goes to all people from the Castle: its stuff, students and of course Assoc. Prof. **Dalibor Štys** – head of the Institute of Physical Biology at University of South Bohemia.

LIST OF ABBREVIATIONS

aa	amino acid residues
AFM	atomic force microscopy
ASU	asymmetric unit
ATP	adenosine-5'-triphosphate
bp	base pairs
CD	circular dichroism
CCD	charge coupled detector
DLS	dynamic light scattering
<i>D</i> _{max}	maximum dimension
dsDNA	double stranded deoxy-ribonucleic acid
EMSA	electrophoretic mobility shift assay
GST	glutathione <i>S</i> -transferase
HJ	Holliday junction
Type I R-M systems	type I restriction-modification systems
MAD	multi-wavelength anomalous dispersion
M.EcoR124I	EcoR124I methyltransferase
nt	nucleotide
REase	restriction endonuclease
<i>R</i> _g	radius of gyration
RIP	radiation induced phasing
SF2	helicase super family 2
SAD	single-wavelength anomalous dispersion
SAdoMet	S-Adenosylmethionine
SANS	small angle neutron scattering
SAXS	small angle X-ray scattering
SeMet	L-Selenomethionine
TFO	triplex forming oligonucleotide
TRD	target recognition domain

TABLE OF CONTENTS

1. SUMMARY.....	1
2. SOUHRN.....	3
I. INTRODUCTION.....	5
3. RESTRICTION-MODIFICATION SYSTEMS. GENERAL CONSIDERATIONS AND CLASSIFICATION.....	5
4. TYPE I RESTRICTION-MODIFICATION SYSTEMS. OVERVIEW.....	8
4.1 Classification and Overall Architecture.....	8
4.2 Activity Control.....	10
4.3 MTase and the DNA Recognition in type I R-M Systems.....	11
4.3.1 HsdS Structure and Action.....	14
4.3.2 HsdM Structure and Action.....	17
4.4 EcoR124I HsdR and its Role in the Complex Functionality.....	18
4.4.1 HsdR. Main Principles	18
4.4.2 DNA Endonuclease Cleavage.....	18
4.4.3 DNA Translocation Blockage.....	21
4.4.4 Bidirectional DNA Translocation. Rates and Kinetics	22
4.4.5 HsdR in the Context of Helicases and Translocases.....	24
4.4.6 HsdR. Structural Information.....	25
5. GOALS OF THE WORK.....	26
II. METHODS.....	27
6. THEORETICAL METHODS. MACROMOLECULAR CRYSTALLIZATION AND STRUCTURE DETERMINATION.....	27
6.1 Crystallization Basics.....	27
6.2 X-ray Crystal Data Collection and the Phase Problem.....	30
6.2.1 Molecular Replacement.....	31
6.2.2 Isomorphous Replacement.....	33
6. Anomalous Dispersion.....	34
7. PRACTICAL METHODS.....	36
7.1 Expression and Purification of HsdR.....	36
7.1.1 Expression and Purification of the Native Recombinant HsdR.....	36
7.1.2 Expression and Purification of the SeMet-labeled Recombinant HsdR..	37
7.3 Crystallization of the Native and Labeled HsdR.....	38
7.4 X- ray Data collection.....	39
7.4.1 X- ray Data collection on Native HsdR Crystals.....	39
7.4.2 X- ray Data Collection on Labeled HsdR Crystals.....	39
7.5 Structure Determination.....	39

7.6 Prediction and Modeling.....	40
III. RESULTS.....	42
8. Expression and Purification of HsdR.....	42
8.1 Expression and Purification of the Native Recombinant HsdR.....	42
8.2 Expression and Purification of the SeMet-labeled Recombinant HsdR.....	43
9. Crystallization of Native and Labeled HsdR.....	44
10. Preliminary X- ray Diffraction Analysis of the Native and Labeled HsdR.....	46
11. STRUCTURE OF THE MOTOR SUBUNIT OF ECOR124I ENZYME.....	48
11.1 Structure Determination and Quality.....	48
11.2 Subunit Architecture: A Planar Array of Functional Domains.....	49
11.2.1 Conservation and Divergence in the Endonuclease Domain.....	51
11.2.2 Functionally Integrated Helicase Domains.....	53
11.2.3 A Helical Domain for Subunit Assembly.....	58
III. DISCUSSION.....	62
12. 180 ROTATION OF HELICASE DOMAIN 2.....	62
13. STAGES OF A ROTARY- INCHWORM STEPPING MOTOR.....	65
14. PENTAMERIC RESTRICTION COMPLEX.....	70
15. DNA TRANSLOCATION AND CLEAVAGE.....	74
CONCLUDING REMARKS AND OTHER PROJECTS.....	78
REFERENCES.....	80
APPENDIX.....	91
List of Presentations.....	91
List of Publications.....	93
CURRICULUM VITAE.....	94
ATTACHMENT.....	96

1. SUMMARY

Type I Restriction-Modification (R-M) enzymes have been developed into highly sophisticated and essential macromolecular complexes acting as a tool for eliminating foreign DNAs during bacteriophage infection. Prokaryotic type I R-M systems effectively recognize and destroy phage DNA by cooperative recruitment of endonuclease, ATPase and DNA translocase and protect the host genome from being self-restricted by DNA methyltransferase action. The complicated interplay of activities performed by distinct components of type I R-M systems could be in many ways translated to the functionally related eucaryotic components of genome remodeling machines, DNA-mismatch repair systems and recombination associated proteins in order to shed the light onto certain aspects of their structural arrangements, functionalities and mode of action.

The central part of this work is an extensive analysis of the crystal structure of the EcoR124I HsdR subunit, which represents the motor subunit of the type I R-M system containing the endonuclease, ATPase and DNA translocase. The structure of 120 kDa HsdR subunit in complex with ATP was solved by Single Wavelength Anomalous Dispersion (SAD) phasing using SeMet X-ray data to 2.6 Å resolution. The resolved structure allowed us to define distinct structural domains, to locate the residues involved in creating a broad protein-protein interaction network between them, and helped to assign catalytic activities.

The highly resolved ATP-binding pocket enabled us to locate residues crucial for ATP coordination, hydrolysis and define similar and distinct features in comparison with other known models of ATPases and translocases. The structures of other representatives of the DEAD-box SF2 family proteins were analyzed to propose a model for HsdR-dsDNA

interaction with its further implementation in describing the coupling mechanism of ATP hydrolysis and DNA translocation.

EcoR124I acts as a pentameric macromolecular assembly. The model of the MTase-DNA complex and various biochemical and biophysical data available for the solved HsdR structure were used to generate a model structure of the entire EcoR124I restriction enzyme complex. For that purpose we used computational techniques to model non-structurally resolved part of the C-terminus of HsdR. The analysis of the MTase-HsdR-dsDNA assembly helped to understand deeply translocation and endonuclease action of type I R-M enzymes. Our structure of the motor subunit of EcoR124I is the first known structure of numerous HsdR homologs of described type I R-M system and enables to model and analyze diverse activities of related DEAD-box proteins.

2. SOUHRN

Restrikčně-modifikační (R-M) enzymy typu I se vyvinuly ve vysoce důmyslné a pro život buňky nezbytné makromolekulární komplexy, které zneškodňují cizorodou DNA při infekci bakteriofágem. Spojením endonukleázové, ATPázové a DNA translokační aktivity v jednom enzymatickém komplexu jsou prokaryotické R-M systémy schopny účinně rozpoznat a zničit fágovou DNA, zatímco chrání vlastní bakteriální genom pomocí methylace DNA. Komplikovaná souhra těchto aktivit, které se účastní jednotlivé komponenty R-M systémů typu I, může být na základě funkčních podobností využita k objasnění strukturního uspořádání, funkce a mechanismu účinku u eukaryotických genom remodelujících faktorů, “DNA-mismatch“ opravných systémů a u proteinů účastnících se rekombinace.

Ve zde popsané práci byla provedena detailní analýza krystalové struktury HsdR podjednotky EcoR124I endonukleázy typu I. Struktura 120 kDa HsdR podjednotky v komplexu s ATP byla řešena s rozlišením 2,6 Å pomocí fázování “single wavelength anomalous dispersion“ (SAD) z dat získaných z rentgenové difrakce krystalu značeného SeMet. Výsledná proteinová struktura nám umožnila definovat jednotlivé strukturní domény a lokalizovat aminokyselinové zbytky účastnící se četných protein-proteinových interakcí a katalytických aktivit.

Vysoké rozlišení struktury vazebné kapsy ATP nám umožnilo lokalizovat aminokyselinové zbytky rozhodující v ATP koordinaci a hydrolýze, a definovat podobné a odlišné vlastnosti ve srovnání s jinými známými modely ATPáz a translokáz. Analyzovali jsme struktury jiných zástupců DEAD-box SF2 rodiny proteinů, abychom

mohli vytvořit model HsdR-dsDNA interakce a uplatnit ho v popisu spojení mechanismů ATP hydrolýzy a DNA translokace.

EcoR124I funguje v pentamerickém makromolekulárním komplexu. Model MTase-DNA komplexu, publikovaná biochemická data a biofyzikální data získaná z vyřešené struktury jsme použili k vytvoření modelové struktury celého komplexu EcoR124I restriktivního enzymu. Za tímto účelem jsme využili náročných výpočetních technik k domodelování strukturně nevyřešené C-terminální části HsdR. Analýza komplexu MTase-HsdR-dsDNA pomůže hlouběji porozumět translokačnímu a endonukleázovému mechanismu u R-M enzymů typu I. Struktura motorové podjednotky enzymu EcoR124I je první známou strukturou ze všech HsdR homologů popsáných u R-M systémů typu I a může umožnit modelování a analýzu řady aktivit u příbuzných DEAD-box proteinů.

I. INTRODUCTION

3. RESTRICTION-MODIFICATION SYSTEMS. GENERAL CONSIDERATIONS AND CLASSIFICATION

Restriction enzymes protect bacteria from infections by viruses, and it is commonly accepted as being their major role in nature, thus they function as the main players constituting microbial immune systems [Murray, 2000].

The phenomenon of restriction, identified first for type I R-M systems, laid the foundations for modern molecular biology, and, eventually, led to the discovery of widely used in DNA cloning techniques and highly commercialized type II restriction enzymes. The classical R-M systems of *Escherichia coli* K-12 and *E. coli* B were first to be discovered by Bertani and Weigle back to 1953 [Bertani and Weigle, 1953]. By using temperate phages (λ and P2) they demonstrated that the restriction endonuclease from *E. coli* K-12 digested the phage λ DNA. It was also shown at that time the protective role of the DNA modification against an endonuclease restriction which prevented the successful propagation of an incoming phage genome in bacteria [Arber and Dussoix, 1962]. The classical type I R-M enzymes of *E. coli* K-12 (EcoKI) and B (EcoBI) were not only the first to be detected but also the first to be purified [Linn and Arber, 1968; Meselson and Yuan, 1968].

Restriction enzymes were found only within prokaryotes and certain viruses, all bacteria and archaea seem to code for them. Restriction enzymes are exceedingly varying in size from the small ones, like PvuII (157 aa), HpyAXII (154 aa), PacI (142 aa), RflFI (136 aa) to the giant HsdR of EcoKI (1188), CjeI (1250 aa), CjeNII (1339 aa) and beyond. According to the REBASE (<http://rebase.neb.com>), over 3,800 activities have been

purified and characterized so far (www.neb.com). Usually for any restriction enzymes in the cell there are one or two accompanying modification enzymes (DNA-methyltransferases) acting to protect the host DNA from cleavage by the restriction enzyme. Modification enzymes recognize specifically the DNA sequence, the same one, as the restriction enzyme that they correspond to, where they introduce a methyl groups on one of the adenine or cytosine bases in DNA strands. Restriction enzymes bind methylated recognition sequences but lack the ability to act upon it. Together, a restriction enzyme and its "cognate" modification enzyme(s) form a restriction-modification (R-M) system.

On the bases of subunit composition, cleavage position, sequence-specificity and cofactor-requirements, there are four well-characterized basic types of restriction systems (I-IV) [Roberts, et al. 2005]. The deterministic features of distinct types of R-M systems are summarized below and in **Table 1**. There are total 3869 restriction enzymes known out of which 91 are of type I , 3762 of type II, 11 of type III and 5 for type IV according to the REBASE statistics.

Table 1. Classification of R-M systems (adapted from Sistla and Rao, 2004)

Feature	Type I	Type II	Type III	Type IV
Structural Subunits	Three different	Two identical	Two different	Two different
Enzyme activity	Endonuclease, methyltransferase ATPase	Endonuclease or methyltransferase	Endonuclease, methyltransferase ATPase	Endonuclease
Biochemical Cofactors for DNA cleavage	ATP, Mg ²⁺ (AdoMet)	Mg ²⁺	ATP, AdoMet, Mg ²⁺	GTPase Mg ²⁺
Methylation	AdoMet, Mg ²⁺	AdoMet	AdoMet, Mg ²⁺	no
Recognition sequence	Asymmetric, bipartite	Usually symmetric	Asymmetric	Bipartite, methylated
Cleavage site	Random, remote from recognition site	At or near recognition site	25-27 bp from recognition site	Between methylated bases at multiple positions
DNA Translocation	Yes	No	Yes	Yes

Type I. Type I R-M systems are multisubunit enzymes that function as a single protein complex. The complex is often composed of two HsdR subunits, two HsdM subunits, and one HsdS subunit. Hsd denotes “host specificity of DNA”. After the non-methylated DNA, containing a sequence recognized by a particular enzyme from type I family, is recognized, the complex, while remaining tightly bound to the recognition site, starts to translocate DNA. Translocation occurs bidirectionally toward the statically bound complex, producing extruding DNA loops which could be easily visualized with the help of the AFM. When translocation is impeded by collision with another type I enzyme or by the topology of the DNA substrate, the enzyme introduces a double-strand cut in the

DNA which could be up to several thousands of base-pairs away from the recognition site [Studier and Bandyopadhyay, 1988; Jindrova et al., 2005]. After the cleavage reaction takes place the complex remains bound to the recognition site [Bickle et al., 1978]. ATP hydrolysis, although, continues long after the cleavage reaction but the enzyme does not turn over [Yuan et al., 1972].

Type II. Type II restriction systems are characterized by having individual restriction enzymes and their accompanying modification enzymes encoded by separate genes. They are usually small, with subunits in the 200-350 amino acids range. Typically recognizing specific DNA sequences, type II restriction enzymes cleave at precise positions within or close to the recognition sequence to produce 5'-phosphates and 3'-hydroxyls - the reason by which they have become highly commercialized. Normally they require divalent Mg^{2+} ions for activity.

Type III. These systems are composed of two genes (*mod* and *res*). *Mod* encodes the protein subunit Mod which specifically recognizes and modifies DNA, with *res* encoding for the restriction-efficient subunit Res. Mod-Res assembly is required for an ATP-dependent restriction. Cleavage is preceded by DNA translocation as with type I R-M systems. The enzymes cut at a specific position, located away from one of the two copies of their recognition sequence. The Mod subunit possesses an independent of the Res DNA-methylase activity and methylates an adenine at its N6 position.

Type IV. The substrate for type IV enzymes is modified DNA, having methylated, hydroxymethylated and glucosyl-hydroxymethylated bases. These systems are composed of one or two genes and recognize sequences which have not been well defined except for EcoKMcrBC, which recognizes two dinucleotides of the general form RmC (R - any purine, followed by a methylated cytosine either m4C or m5C) and which are separated

by anywhere from 40-3000 bases. They cut at sites approximately 30 bp away from one of the recognition sequences.

4. TYPE I RESTRICTION-MODIFICATION SYSTEMS. OVERVIEW

4.1 Classification and Overall Architecture

Type I R-M systems are divided into four families based on allelic complementation, protein homologies and biochemical properties: types IA (including EcoKI and EcoBI), IB (EcoAI) and ID R-M systems which are chromosomally encoded [Barcus et al., 1995; Titheradge et al., 1996] while most type IC (typified by EcoR124I) systems are carried on large conjugative plasmids [Firman et al., 1983]. It was also proposed that KpnBI system discovered in *Klebsiella pneumoniae* and which does not share enough homology with any preexisting type I family could represent the prototype of a new type, designated as type IE [Chin et al., 2004]. Below, the focus will be mostly on the information coming from the experiments on type IC EcoR124I enzyme, however, pronounced differences in comparison to enzymes from other families will be also highlighted.

The type I R-M EcoR124I enzyme has been discovered during the analysis of the R124 plasmid of *E. coli*, which is a large conjugative plasmid encoding a DNA restriction and modification system [Hedges and Datta, 1972; Bannister and Glover, 1968]. The plasmid carries two specificities R124 and R124/3 with only one of them is normally expressed, and it is supposed that the one specificity has emerged from the gene duplication and sequence divergence of the other and activities are able to be switched under certain conditions [Firman et al., 1983; Glover et al., 1983].

Type I R-M enzymes are large, multifunctional macromolecular complexes composed of three different subunits: HsdS, HsdM and HsdR [McClelland and Szczelkun, 2004].

Subunits are coded by the corresponding *hsd* genes, for which, transcription of *hsdM* and *hsdS* happen from the same promoter (Pmod), while *hsdR* has its own promoter (Pres) [Murray, 2000]. The activities of the complex of all three subunits include ATP-dependent DNA translocation, DNA cleavage and methylation [Szczelkun et al., 1996; Firman and Szczelkun, 2000; Seidel et al., 2004; McClelland et al., 2005]. HsdM, homing a binding site for SAdoMet and catalytic residues for the methylation, together with HsdS form an active methyltransferase (MTase) with stoichiometry M_2S_1 , which, sufficient for transferring methyl groups on DNA [Webb and Taylor et al., 1995]. The MTases from different types of R-M enzymes differ in their size. For example, the EcoR1241 methyltransferase (M.EcoR1241) has been shown to consist of two copies of the HsdM subunit (each 58 kDa) and one HsdS subunit (46 kDa), to form a trimeric enzyme (162 kDa) [Taylor et al., 1994]. The observed molecular mass of the endonuclease according to gel filtration analysis suggests that EcoR124I has a stoichiometry of $R_1M_2S_1$ (calculated $M_r = 281.8$ kDa) [Janscak et al., 1996], unlike those of EcoKI which carries two HsdR subunits [Weiserova et al., 1993], bound to the MTase with similar affinities [Dryden et al., 1997]. In later experiments employing an EMSA it was found that the purified EcoR124I enzyme was actually a mixture of two species with larger species having a stoichiometry of $R_2M_2S_1$ and the smaller is $R_1M_2S_1$ [Janscak et al., 1998] but the $R_2M_2S_1$ complex was able to dissociate onto the $R_1M_2S_1$ complex and the HsdR subunit with an apparent K_d of $\sim 2.4 \times 10^{-7}$ M displaying different interaction affinities of HsdR with the MTase in the assembly of the EcoR124I [Janscak et al., 1998]. M_2S_1 is the only one form observed for EcoR124I MTase and no M_1S_1 complex has been detected [Taylor et al., 1992].

The purified EcoBI, on the other hand, exists in a number of forms $R_2M_2S_1$, $R_1M_2S_1$ and $R_1M_1S_1$ [Eskin and Linn, 1972]. EcoKI is a stable $R_2M_2S_1$ complex but separately from HsdR, EcoKI MTase (M_2S_1), however, dissociates into an inactive M_1S_1 species and free HsdM subunit [Dryden et al., 1993; Dryden et al., 1997]. EcoAI forms a weak complex that dissociates into MTase and HsdR subunits when purified [Suri et al., 1984].

4.2 Activity Control

Binding of the first EcoR124I HsdR subunit to the MTase is much stronger than binding of the second HsdR to the $R_1M_2S_1$ complex and the intermediate $R_1M_2S_1$ complex is not able to cleave DNA [Janscak et al., 1998]. Above facts are proposed to be the core of the sensitive regulation mechanism of restriction activity following conjugative transfer of the EcoR124 *hsd* genes into a non-modified recipient cell [Janscak et al., 1998]. After the conjugation transfer, tight M_2S_1 complex is assembled first in the recipient cell. This complex further traps available HsdRs in a cleavage inefficient but methylase proficient $R_1M_2S_1$ complex, which modifies the recipients DNA. Accumulation of the HsdR on the later stages establishes the fully operational R-M system in a new host [Janscak et al., 1998; Janscak et al., 1996]. Relatively high methylation activity of EcoR124I with non-modified DNA supports the above regulation mechanism ensuring a rapid modification of a non-modified chromosome in the recipient cell following conjugal transfer of the *hsd* genes [Janscak et al., 1998; Janscak et al., 1996].

Surprisingly, the cellular localization of different type I families also varies, which could be related to the control of activity of these enzymes *in vivo*. Although all enzymes interact with the cytoplasmic membrane, the nature of these interactions differs from family to family. EcoKI, for example, is associated with the inner membrane likely via

the interaction with chromosomal DNA, with the HsdR subunit of the complex being exposed on the outer surface of the cytoplasmic membrane [Holubova et al., 2000]. Similar localization have been reported for the EcoR124I enzyme, while the HsdR subunit of EcoAI must be anchored within the spheroplast membrane and resisted the protease action on it at the periplasmic surface of the inner membrane [Holubova et al., 2004].

4.3 MTase and the DNA Recognition in Type I R-M Systems

The MTase, consisting of HsdM and HsdS subunits with the composition M_2S is responsible for the specific DNA sequence recognition and together with the HsdR subunit forms a functional endonuclease [Bickle, 1993]. The DNA sequences recognized by different type I R-M systems have been determined [Price et al., 1978]. The cognate recognition site for the EcoR124I HsdS is 5'GAANNNNNNRTCG 3' sequence [Price et al., 1978], where N is any base and R is a purine. The N6 of adenine base on each strand of DNA is the site of methylation [Taylor et al., 1993]. The recognition sequence of EcoR124II differs by one additional base in the nonspecific region [Price et al., 1978]. HsdS imparts to the enzyme a bipartite asymmetric DNA sequence specificity binding, recognizing two specific DNA stretches of 3 or 4 bp one and the other of 4 or 5 bp, interrupted by a 6 to 8 (depending on the system) nucleotides long nonspecific spacer [Roberts et al., 2003; Price et al., 1978].

Substrate recognition and selectivity in the type IC DNA modification methylase M.EcoR124I studied with the use of exonuclease III digestion footprint technique showed that M.EcoR124I specifically binds to the DNA sequence and protects a 25 bp fragment including its cognate recognition site [Taylor et al., 1993]. EcoR124I MTase tightly

encloses the DNA in the complex, protecting ~23 bases on each strand in the DNase I footprint, in concord with the pronounced hydroxyl radical footprinting and SAdoMet had no significant effect on the extend of the footprint [Mernagh and Kneale, 1996]. Adding the HsdR to the M.EcoRI24I does not increase the size of the footprint but increases the stability of the resulting DNA-protein complex [Mernagh et al., 1998]. The lack of the extended footprint, was probably due to the absence of ATP analogues in the footprint studies, which is needed for a small DNA bulge formed by the initial R1 or R2 complex, as was shown in AFM experiments [van Noort et al., 2004]. The bulge could be a ssDNA as was shown by experiments involving single-strand nuclease P1 with the complementary strand being protected by the strong interaction with the enzyme complex. The bulge formation, accompanied by a shortening of the contour length of the DNA molecule, represents the first-step of the initial loop formation [van Noort et al., 2004].

Structural features of the EcoRI24I MTase and its behavior upon the binding of a cognate DNA substrate in different methylation states began to emerge after the work of Taylor and colleagues [Taylor et al., 1994]. Small angle X-ray solution scattering (SAXS) measurements along with CD revealed an unusually large compaction of the EcoRI24I MTase upon its binding to a cognate recognition site on oligonucleotide duplex, seeing in the reduction of nearly 70 Å in the overall dimensions of the complex and may be explained by structural change involving a major rearrangement of the subunits. The above experiments led to the proposition of the model showing the open and closed forms of M.EcoRI24I corresponding to the free enzyme and the complex with DNA [Taylor et al 1994]. The big conformational changes upon substrate binding has been also confirmed by the probing the domain structure of type IC DNA methyltransferase EcoRI24I by

limited proteolysis, in which the number of protease accessible sites have been decreased for the substrate bound MTase comparing to the free form [Webb et al. 1995]. Binding of SAdoMet or its analogues showed no significant effect on the digestion of EcoR124I MTase [Webb et al., 1995]. Interestingly, contrast results were demonstrated for the M.EcoKI. Unlike for M.EcoR124I, M.EcoKI does not show a decrease of the degree of the proteolytic protection upon the substrate binding [Cooper and Dryden, 1994], and does so upon interacting with the SAdoMet [Powell et al., 1993]. Lysine residues of M.EcoR124I are crucial for the DNA interaction as has been shown using chemical modifications of the surface accessible lysines in the free enzyme and in its complex with cognate DNA sequence. A total of 41 of 109 lysine residues in the enzyme are susceptible to modification, of which 19 of HsdS and 22 of both HsdM subunits and the lysine residues in the HsdM subunit are protected for modification to a less extent than those in HsdS [Taylor et al., 1996]. Large conformational changes in M.EcoR14I upon the DNA binding are also accompanied by considerable distortion of the DNA structure [Taylor et al., 1994].

No crystal structure has been solved for the trimeric MTase for which, however, a low-resolution structure of the AhdI MTase, from *Aeromonas hydrophila* determined by SANS is available, that reveals the overall shape and subunits organization of the complex [Callow et al., 2007]. There is also a structural model for the methyltransferase M.EcoR124I in complex with DNA obtained by docking models of individual subunits generated by fold-recognition and comparative modeling, followed by optimization of inter-subunit contacts by energy minimization [Obarska et al., 2006].

In the SANS model the overall shape of the S subunit dimer resembles crystal structure of HsdS while the M subunits have an extended outer regions which account for the main

difference from the crystal structure, where that outer domains are extremely flexible and not good resolved or missing at all, being probably the subject to considerable crystal packing effects. These extended regions therefore seem to be flexible and wrap around the DNA to form a more globular structure in the MTase-DNA complex [Callow et al., 2007]. In the MTase model proposed by Obarska et al., 2006, flexible domains were placed in an arbitrary position with respect to the DNA bound catalytic domain that does not produce any steric and structural problems. The unknown mutual orientation of the domains indicates that those position in the crystal structure of of the HsdM subunit of EcoKI probably irrelevant to the function, due to the absence of a stabilizing effect of HsdS [Obarska et al., 2006].

4.3.1 HsdS Structure and Action

In EcoR124I the bipartite DNA specificity is realized by the two regions of 150-180 amino acid residues in HsdS, which are highly variable even between members of the same family [Suri and Bickle, 1985]. The two variable regions of HsdS subunit form independent target recognition domains (TRDs). Apart of the variable regions, two conserved regions were identified: a central conserved region and a C-terminus one [Cowan et al., 1999]. The conserved regions have been proposed to be important in the assembly of the functional Mtase, where deletion of the central conserved region impairs the HsdS-HsdM interaction [Abadieva et al., 1994]. HsdS of EcoR124I alone is insoluble, but soluble form was obtained by purifying the GST fusion protein [Mernagh et al., 1997]. Trp212 substituted to Arg at the border of the central conserved region and the second target recognition domain (TRD2) in HsdS, significantly influences the ability of the HsdS to assemble with HsdM and as a consequence of this, the mutant MTase has

drastically reduced DNA binding. The phenotype of such a mutant is Res⁻Mod⁺ and it is not able to cleave DNA [Weiserova and Firman, 1998], but binding of the HsdR restores the methylation activity of the complex [Weiserova et al., 2000].

There are two crystal structures of the specificity subunit of type I enzymes. S.MgeORF438P from *Mycoplasma genitalium* solved at 2.3 Å (PDB ID 1ydx) [Calisto et al., 2005] and S.MjaORF132P from *Methanocaldococcus jannaschii* solved at 2.4 Å (PDB ID 1yf2) [Kim et al., 2005]. The first structure of HsdS came from the SAD data on the phenyl mercury soaked specificity subunit from *M. jannaschii* at 2.4-Å resolution [Kim et al., 2005]. The second structure from *M. jannaschii* shows similar organization. The structures are made of two globular domains, separated by a pair of antiparallel α -helices. The globular domains correspond to the target recognition domains (TRDs) variable in sequence, as previously defined for S subunits on sequence analysis, while the two helices correspond to the central (CR1) and C-terminal (CR2) conserved regions, respectively [Calisto et al., 2005] (**Fig. 4-1**). Based on the *M. jannaschii* HsdS, the model of the DNA complex with the two TRDs in S subunit was obtained and it was shown that the complex structure required a kinking of the dsDNA between the two binding sites, which are spaced by 8 bp and causes unwinding of dsDNA to expose the base to be modified [Kim et al 2005] as was also exploited in the MTase-DNA model by Obarska et al., 2006.

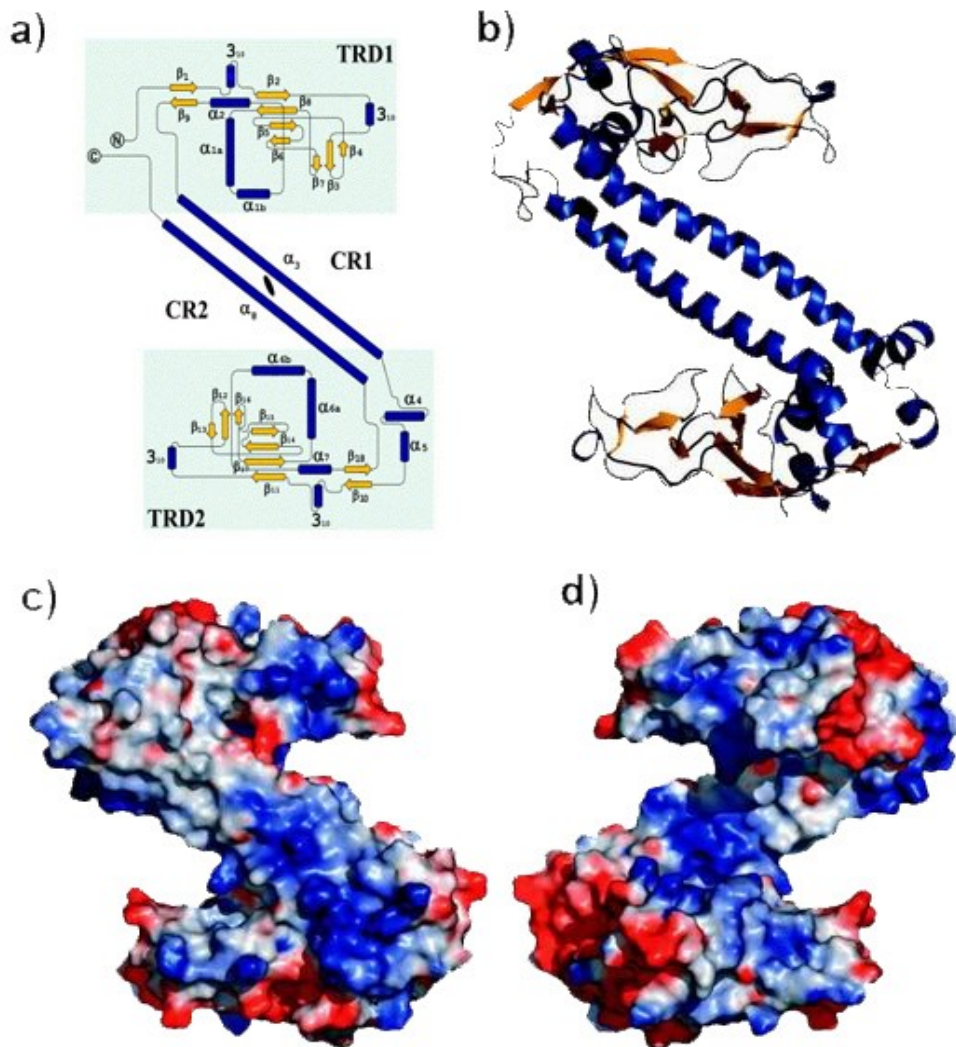


Fig. 4-1 Structure of the HsdS specificity subunit from *Mycoplasma genitalium*. **(a)** Topology diagram of the S subunit of a type I R-M system from *Mycoplasma genitalium*. **(b)** Tertiary structure representation of an S subunit from *Mycoplasma genitalium*. A monomer made of four successive structural domains: the globular TRD1, a long helical CR1 domain, the globular TRD2, and a C-terminal CR2 helix. Two almost identical TRDs are attached to both ends of the central complementary CR helices. **(c,d)** Electrostatic potential surfaces of the S **(c)** and Z **(d)** faces of HsdS. The Z, DNA – binding face, with more pronounced positively and negatively (red) charged residues [adopted from Callisto et al., 2005].

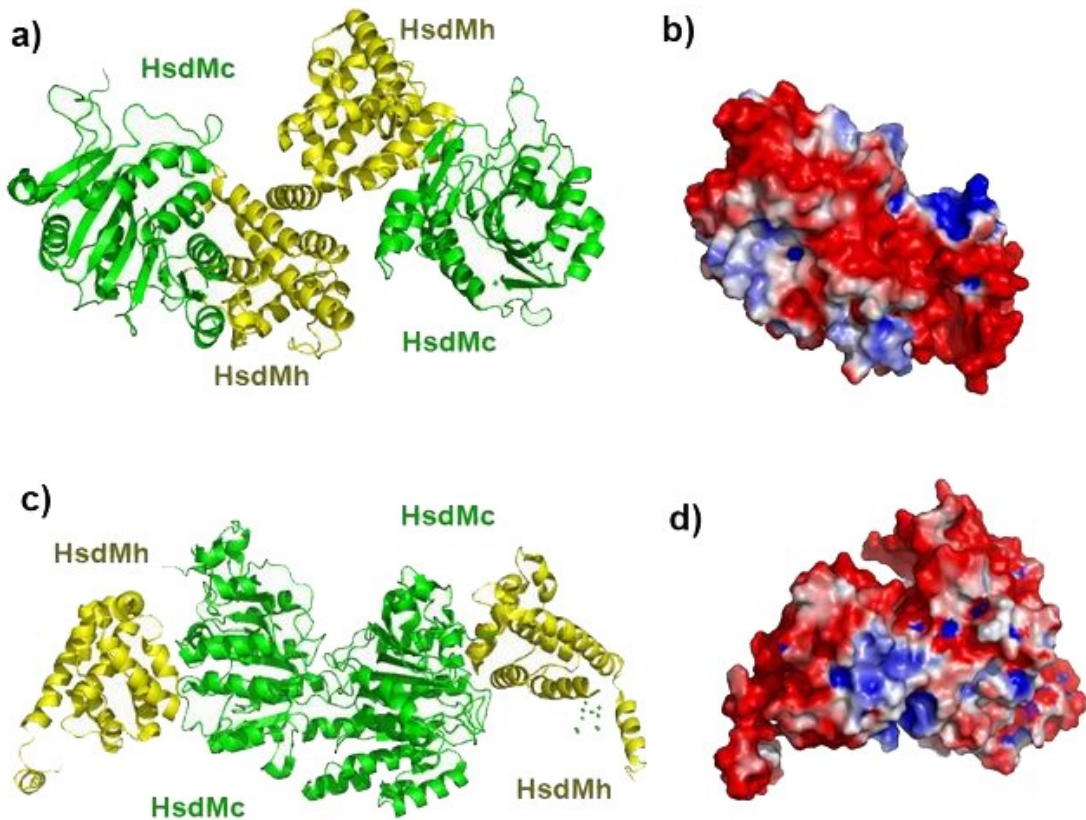


Fig. 4-2 Structures of the HsdM subunits of type I R-M system. Two structures are currently known: **(a)** M.BthVORF4518P from *Bacteroides thetaiotaomicron* solved at 2.2 Å (2okc, unpublished) and **(c)** HsdM of EcoKI from *E. coli* at 2.8 Å (2ar0, unpublished), Structures are represented in the cartoon representation mode. Both structures make a dimer in the ASU. Each monomer comprises two domains: HsdMc – catalytic domain, containing residues which bind the cofactor for DNA methylation – SAdoMet; and the HsdMh – the whole helical domain, not fully resolved in EcoKI crystal structure. Whereas in the dimer from *B. thetaiotaomicron* two monomers interact via their HsdMh domains, in the structure from *E. coli* the monomers contacts are through HsdMc only. **(b,d)** Electrostatic potential surfaces of EcoR124I HsdM protein monomers with orientation as in corresponding dimers on the left [unpublished data].

4.3.2 HsdM Structure and Action

The HsdM subunit homes the catalytic site for DNA methylation as well as the binding site for the methyl donor and restriction cofactor SAdoMet [Willcock et al., 1994]. Two structures are currently known: M.BthVORF4518P from *Bacteroides thetaiotaomicron*

solved at 2.2 Å (PDB ID 2okc, unpublished) and HsdM of EcoKI from *E. coli* at 2.8 Å (PDB ID 2ar0, unpublished) (**Fig. 4-2**).

4.4 EcoR124I HsdR and its Role in the Complex Functionality

4.4.1 HsdR. Main Principles

HsdR subunit is the biggest part of the complex and is responsible for the endonuclease DNA cleavage and it is also an ATP-dependent DNA translocase. Being an important part of the complex, HsdR subunit started to be thoroughly analyzed to unravel unexpected properties of the subunit [Zinkevich et al., 1997]. The molecular weight of HsdR from EcoR124I is 120 kDa and it is composed of 1038 residues. The subunit domain architecture, including distribution of conserved motifs is schematically drawn on **Fig. 4-3**. Gel filtration of HsdR demonstrated that the subunit exists in a monomeric form in solution [Zinkevich et al., 1997]. DLS, Sedimentation velocity experiments and SANS also supported that HsdR exists as a monomer and does not complexate in solution [Obarska-Kosinska et al., 2007]. The optimization of the production of EcoR124I enzyme made it available to numerous *in vivo* studies, resulting in deeper analysis and discovering a new biochemical features [Janscak et al., 1996].

4.4.2 DNA Endonuclease Cleavage

Alone, the HsdR subunit was capable to hydrolyze ATP but very slow in comparison to the reconstituted enzyme and was inhibited when specific DNA was present. The technique of EMSA showed the inability of HsdR to bind oligoduplexes carrying the EcoR124I recognition site, although, some non-specific fragments were found to be retarded with a fairly weak constants of 0.2 [μ]M [Zinkevich et al., 1997]. Mg^{2+}

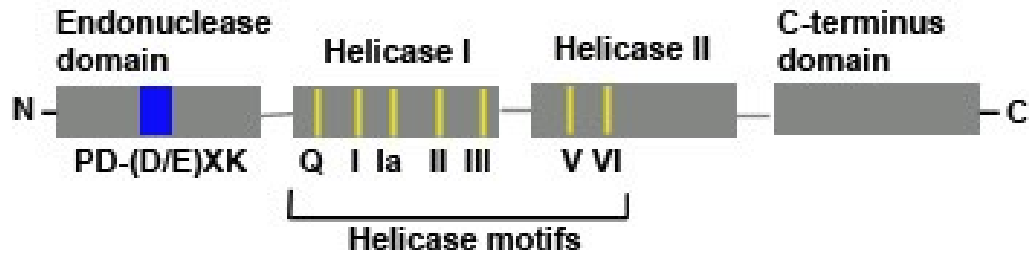


Fig. 4-3 Subunit domain architecture of the ATP-dependent restriction endonucleases HsdR subunit. Protein domains are represented as gray rectangles approximately to scale. Blue rectangle represents the Region X, comprised of the conserved residues which build up the endonuclease cleavage site. Yellow vertical lines indicate a collection of helicase motifs distributed over the two helicase domains. All helical C-terminus domain is adjacent to the helicase domain 2 [McClelland and Szczelkun, 2004].

-dependent non-specific nuclease activity was also observed [Zinkevich et al., 1997]. According to the model, explaining the distant location of the DNA cut from the cognate sequence by the event of the DNA tracking by the enzyme [Studier and Bandyopadhyay, 1988; Bickle, 1993], requirements for cleavage were investigated. Cleavage of two-recognition-site linear DNA by EcoR124I produced sets of heterogeneous products expected to be generated by random distant cleavage between the two EcoR124I sites that concur with the collision model proposed by Studier and Bandyopadhyay, 1988. Linear DNA molecules containing a single recognition site are either refractory to cleavage or undergo limited cleavage at a high excess of enzyme over DNA [Szczelkun and Dillingham 1996; Janscak et al., 1999a]. In case of the circular one-site plasmids random cleavages throughout the DNA were observed. The cleavage may result from topological constraints of the DNA concurrent with translocation by a single complex [Szczelkun et al., 1997]. In contrast to the remote cleaved products, persistent cleavages at loci close to the EcoR124I recognition sites (within 250 bp) were observed with no specificity to any particular DNA [Szczelkun et al., 1997]. Based on the above experimental data the model of nicking and cleavage of expanding DNA loops by

EcoR124I was proposed [Szczelkun et al., 1997], in which an excessive build up of supercoiling torque in extruding loops during the DNA motion could be overcome by the generation of nicking events in loops, accounting for the hydrolysis close to the recognition site [Szczelkun et al., 1996]. The influence of the degree of negative DNA supercoiling on the cleavage was investigated, showing that the increase in the degree of negative supercoiling, lowers the cleavage efficiency, analyzed by the accumulation of the linear form of DNA [Janscak et al., 1996].

An ubiquitous type I R-M enzymes cofactor SAdoMet appeared to be not required by R.EcoR124I, which, however, stimulates the rate of ATPase activity and DNA cleavage [Janscak et al., 1996], elevating the binding affinity to the cognate sequence as seen from the surface plasmon resonance analysis [Janscak et al., 1996].

A lot has become known about the catalytic properties of the whole EcoR124I complex, mode of cleavage and the nature of DNA. Continuing on the DNA cleavage mechanism of type I restriction enzymes the nature of the DNA molecule termini produced in the cleavage reaction by types IA, IB and IC was investigated [Jindrova et al., 2005]. Data indicated that these enzymes can generate both 5'- and 3'-overhangs of various lengths, whereas blunt ends were not detected [Jindrova et al., 2005]. Whereas, the preferences for the short overhang termini of 2-7 nt formation was observed, the formation of long 3'-overhangs also occurred, in concord with the known fact of 3'-overhangs of 100 nt long described for the type IA restriction enzyme EcoBI [Kimball and Linn, 1976; Endlich and Linn, 1985]. The HsdR subunit contains only a single catalytic set of residues for phosphodiester bond cleavage, so called Region X, which is located before the DEAD-box motifs in the primary sequence and shares similarities with the PD-(D/E)XK motif found in many type II REases and a range of other nucleases involved in

DNA repair and recombination [Aravind et al., 2000; Bujnicki and Rychlewski, 2001; Kosinski et al., 2005]. Mutations in the key residues of the Region X abolished the type IB enzyme EcoAI restriction activity but had no effect on the directional DNA motion [Janscak et al., 1999b]. To introduce a double strand cleavage, two nuclease catalytic centers must be brought to close proximity on the DNA: one from the translocating HsdR subunit and the other either from the HsdR belonging to converged type I enzyme or by HsdR subunit coming from the solution [Jindrova et al., 2005].

4.4.3 DNA Translocation Blockage

The DNA translocation blockage as a general mechanism of cleavage site selection by type I restriction enzymes has been further experimentally probed [Janscak et al., 1999]. The experiments with potential blocks to the translocation being introduced into the DNA, such as positive supercoils, Holliday junctions (HJ), or collided enzymes from different families, indicated that irrespective of the topological status of DNA, type I restriction enzymes cleavage is triggered after the translocation process is halted [Janscak et al., 1999a]. Data, however, suggests that positive supercoiling does not act as the barrier to DNA translocation that leads to DNA cleavage. Experiments in which a supercoiled DNA domain was linked by a HJ evidenced that cleavage of circular DNA is not triggered by accumulation of torsional stress in the contracting DNA loop. EcoR124I persistently cleaved one on either side of the junction. If the translocation process was blocked by changes in DNA topology, the enzyme could have not reached the HJ, located 1240 bp in one direction and 1110 bp in the other, from the recognition site and DNA cleavage would be observed within the regions between the junction and the enzyme

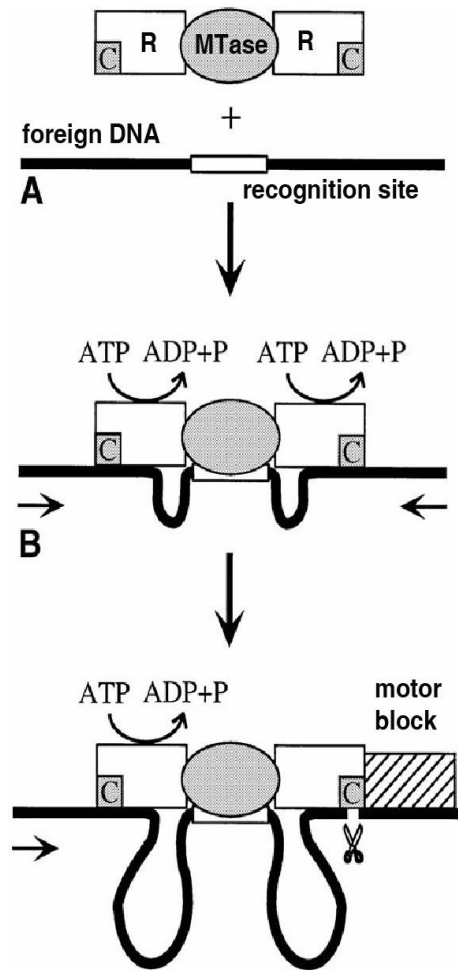


Fig. 4-4 A model for DNA cleavage triggering due to translocation blockage. The HsdR subunits are shown on both sides of the MTase. Foreign DNA with the recognition sequence and the motor block are schematically represented. [Janscak et al., 1999a]

recognition site [Janscak et al., 1999a]. Experiments, stated that DNA translocation blockage is the only requirement for DNA cleavage to take place, leading to the model of DNA cleavage by type I restriction enzymes [Janscak et al., 1999a] (**Fig. 4-4**). In addition to restriction of foreign DNA, type I restriction enzymes may participate in special types of recombination of foreign DNAs [Kusano et al., 1997; Janscak et al., 1999], although the migration of the HJ into the heterologous DNA was ruled out [Stanley and Szczelkun, 2006]. Biochemical studies indicated that the integrity of the 5'-3' strand is more important for activity, while gaps in the 3'-5' strand can be tolerated during translocation. Recent single molecule and bulk solution measurements on the dsDNA motor protein EcoR124I indicate that an intact 5'-3' strand is required for translocation, whereas the 3'-5' strand only assists in processivity [Stanley et al., 2006].

4.4.4 Bidirectional DNA Translocation. Rates and Kinetics

It was noted that the dynamics of the nucleic acid motors remain inadequately characterized. Such fundamental questions like what is the translocation rate, direction of movement relative to binding-site polarity, and how often does an enzyme stall on, or

release the DNA were addressed in the work by Firman and Szczelkun by evaluating protein motion on DNA by the type I restriction endonuclease EcoR124I using TFO displacement [Firman and Szczelkun, 2000]. The direct *in vitro* assay of protein allowed direct measurement of translocation velocity at a speed of 400 bp/s at 20°C indicating the high processive DNA translocation by EcoR124I. R₁M₂S₁ complex cannot cleave DNA [Janscak et al., 1998] but still translocates it, however, with less processivity than that catalyzed by the R₂M₂S₁ form, as events of dissociation from the DNA during motion are more frequent. Therefore, the presence of two HsdR subunits stabilizes the translocating complex such that dissociation is rare [Firman and Szczelkun, 2000]. Later the fluorescent triplex assays and single molecule studies using EcoR124I allowed continuous monitoring of DNA translocation [McClelland et al., 2005] in which the rate of translocation and a rate of initiation were quantified [McClelland et al., 2005; Seidel et al., 2004]. Single molecule assay showed that the translocation rate is independent of DNA sequence and is largely invariant within the enzyme population (550(±90) bp/s at 25 °C with the average distance translocated by EcoR124I before a dissociation event being >3400 bp [Seidel et al., 2004]. In order to get deeper insight into the biochemical mechanism of DNA translocation, a detailed characterization of the ATPase activity of EcoR124I holoenzyme was carried out utilizing a coupled spectrophotometric ATPase assay to study ATP hydrolysis on either linear or supercoiled dsDNA containing a single recognition sequence for EcoR124I holoenzyme, showing that 3.1(±0.4) ATP molecules were hydrolyzed per base-pair translocated per R subunit. The poor efficiency was explained by transiently disengaging HsdR subunit from DNA on average, every 515 bp during translocation, while the MTase core unit stays bound at the DNA [Bianco and Hurley, 2005; Seidel et al., 2005]. Recently, the understanding of the dsDNA

translocation by EcoR124I enzyme has been extended to include the ATP coupling efficiency and motor step size [Seidel et al., 2008]. The bidirectional motion of the true dsDNA translocase EcoR124I along dsDNA was calculated to take steps of 1–2 bp, with ~1 ATP consumed for each base pair moved. The data led to the generation of the models for different stepping modes of EcoR124I [Seidel et al., 2008].

During translocation there is no detectable strand separation by EcoR124I, as it was shown using intrinsic fluorescence property of the *E. coli* SSB protein as a reporter for dsDNA unwinding [Bianco and Hurley, 2005]. It was found that EcoR124I can translocate past covalent interstrand crosslinks also excluding the unwinding activity [Stanley et al., 2006].

4.4.5 HsdR in the Context of Helicases and Translocases

Helicases are nucleic acid-dependent ATPases that unwind DNA or RNA duplex substrates. As a consequence, they participate in almost every process in cells that recruits nucleic acids, including DNA transcription, splicing, transport, translation, and degradation of mRNA and ribosome biogenesis [Rocak and Linder, 2004]. A wide variety of proteins from bacteria to man can be classified on the basis of characteristic amino acid motifs into one of four helicase superfamilies –SF1, SF2, SF3 or SF4 (Hall and Matson, 1999). Type I restriction enzymes fall into the SF2 family featured by the centrally located collection of helicase motifs (DEAD-box motifs) (**Fig. 4-3**), making them similar to the so-called 'DEAD box' proteins, and associated with the ATP-dependent translocation of DNA that precedes restriction [Murray et al., 1993; Gorbalenya and Koonin, 1991]. The crystal structures of a various of SF1 and SF2 DNA and RNA helicases revealed that the superfamily motifs cluster into an ATP-binding

pocket in between two RecA-like domains [Singleton and Wigley, 2002; Caruthers and McKay, 2002]. The central core of the pocket comprises Motifs I (Walker A) and II (Walker B) and Motif VI (Arginine fingers). These motifs coordinate the magnesium ion and phosphate groups of Mg-ATP complex and are more conserved along a number of other helicases and translocases. Mutations at some of the key residues inside the motifs affect the coupling of ATP hydrolysis to DNA motion and cleavage [Webb et al., 1996; Davies et al., 1999b). Other motifs (Q, Ia, III, Y, V) are involved in inter domain interactions and binding of the nucleic acids [McClelland and Szczelkun, 2004].

4.4.6 HsdR. Structural Information

Out of 28 restriction enzymes, structures of which have been solved and deposited, all belong to the type II R-M system (REBASE statistics). The first model of the entire HsdR was generated based on the SANS data constraints (R_g of 3.4 nm and a D_{max} of 10 nm) and homology modeling based on the known structures of DNA helicases and endonuclease NgoMIV [Obarska-Kosinska et al., 2007]. However, HsdR of EcoR124I has been successfully crystallized by us and the preliminary diffraction analysis showed a resolution up to 2.6 Å [Lapkouski et al., 2007]. The structure was solved and we deposited the structure at 2.6 Å resolution to the Protein Data Bank (PDB ID 2w00).

5. GOALS OF THE WORK

EcoR124 I is an enzyme, belonging to the type I Restriction-Modification (R-M) system, which is a bacterial cell defender against bacteriophages. Phages attach to the bacterial cell wall inserting a genome material into the cytoplasm. They intend to exploit the cell replication and protein synthesis machinery to replicate and propagate, what eventually can lead to bacteria lysis. Type I R-M enzymes have been evolved as multisubunit complexes, homing different types of activities, specifically designated toward an effective elimination of the foreign DNA by endonuclease cleavage, done by HsdR subunits.

The goals of the Ph.D work are:

- 1) To express and purify the recombinant native and SeMet labeled (derivative) HsdR subunit of EcoR124I R-M complex;
- 2) To find optimal crystallization conditions for obtaining quality native and derivative HsdR crystals, suitable for crystallography experiments;
- 3) To collect and analyze the diffraction data from the HsdR crystals and to determine the atomic structure of the HsdR protein;
- 4) To perform a detailed structural analysis of the HsdR structure, identifying functional domains; residues involved in ATP hydrolysis, interdomain interactions and DNA binding;
- 5) Based on known structural and biochemical data, to generate the model of the entire EcoR124I complex to study the gross structural features of the enzyme;
- 6) To produce a reliable and testable molecular mechanism of coupling of ATP-dependent DNA translocation and cleavage activities, which could be applied to other type I restriction-modification systems.

II. METHODS

6. THEORETICAL METHODS. MACROMOLECULAR CRYSTALLIZATION AND STRUCTURE DETERMINATION

The most common method of obtaining a detailed model of a protein molecule, allowing the resolution of individual atoms, is to interpret the diffraction of X-rays from electrons of many identical molecules in an ordered crystalline array - single-crystal X-ray crystallography. X-rays are electromagnetic radiation, diffracted by even the smallest molecules. The wavelength of X-rays, roughly spoken, should not be larger than the molecules themselves, what is a prerequisite for production an image of individual atoms in protein molecules using a diffraction experiment. It is, however, still not possible to produce a focused image of a single molecule, because, X-rays cannot be focused by lenses what is bypassed by measuring the directions and strengths (intensities) of the diffracted X-rays and then using a computational techniques to simulate a lens that reconstructs an image of the macromolecule.

6.1 Crystallization Basics

In order to determine the structure of a protein by the way of X-ray crystallography high-quality crystals (with edges around 0.1-0.3 mm) of the purified protein is unavoidable requirement. The one of the most distributed and reliable procedure of growing protein crystals is to mix a purified protein with a precipitant such as ammonium sulfate or polyethylene glycol, at a concentration of the precipitant just below that necessary to precipitate the protein. Then the system is subjected to the slow controlled evaporation process in which, water is removed to raise protein concentration, leading to the crystal

formation through two stages of nucleation and growth. The essence, of what is going on during the crystallization could be grasped from the the phase diagram, which is a map representing the state of a material (e.g., solid and liquid) as a function of the ambient conditions (e.g., protein and precipitant concentration) [Asherie, 2004] (**Fig. 6-1a**). Crystals can only form in supersaturated solutions, so an ideal strategy (**Fig. 6-1b**) would be to start with conditions corresponding to the blue region of the phase diagram, and then, when nuclei (tiny molecular clusters giving birth to crystals) form, move into the green region, where growth, but not additional nucleation, can occur (**Fig. 6-1b**) [Rhodes, 2006]. There are several techniques for setting up a crystallization trial: sitting and hanging drop vapor diffusion, batch, sandwich drop, microbatch under oil, microdialysis and counter diffusion. The mostly used crystallization techniques are vapor diffusion methods using sitting or hanging drop crystallization, performed by mixing the droplet of sample with crystallization reagent on a platform, located as an elevation over the reservoir in a closed container. The initial lower concentration of the reagent in drop than in the reservoir accounts for water being pulled away from the droplet leading to the

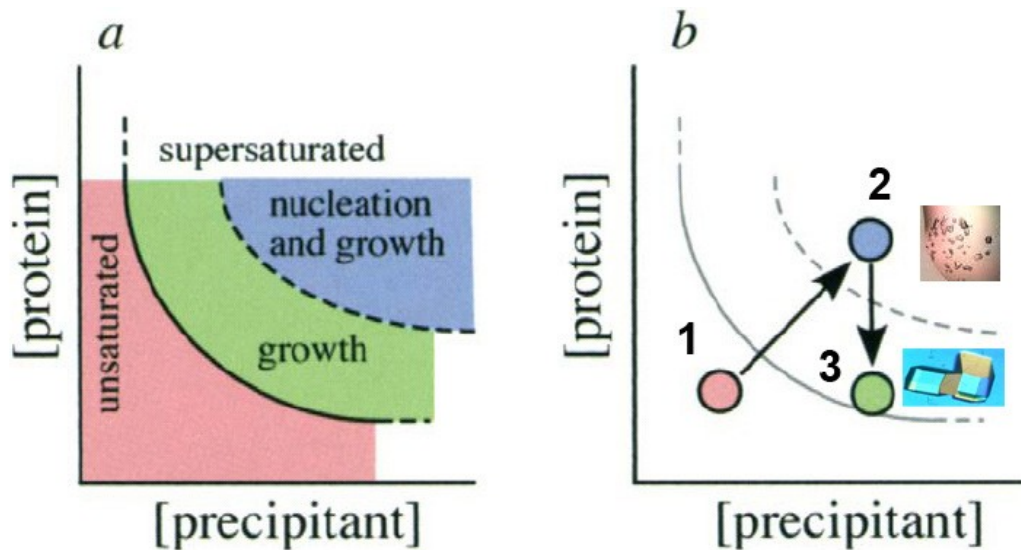


Fig. 6-1 (a) Phase diagram showing the solubility of a protein in solution as a function of the concentration of the precipitant present. The red region represents the conditions under which the protein solution is unsaturated where neither nucleation nor growth occurs. The green and blue regions represent supersaturated solution where in the blue region both nucleation and growth are supported, while conditions in the green support growth only. **(b)** The strategy for growing large crystals for a diffraction experiment is to follow the indicated path, thus allowing nucleation to occur under conditions in the blue region, then slowly move to conditions in the green region [Adapted from Rhodes, 2006].

protein concentration as system evolves toward an equilibrium. The fact that the sitting drop method is easy to perform, requires small amount of sample and a lot of flexibility over screening and optimization, commercial crystallization robots have been modified for use in setting up sitting-drop vapor-diffusion crystallization experiments, allowing very effective high throughput crystallization using nanoliters of sample. The hanging drop technique is also very popular and differs from the sitting drop experimental setup in that the drop with the sample mixed with the precipitant on a cover slide is invertly positioned over the reservoir and as in the sitting drop, the difference in the concentrations of the reagent in the drop and reservoir drives the system toward the equilibrium via the vapor phase. During this equilibration process the sample becomes more concentrated due to the water being pulled away from the less concentrated droplet

solution. Quality data collection for macromolecular cryocrystallography requires suppressing the formation of crystalline or microcrystalline ice that may result from flash-freezing crystals [Rubinson et al., 2000] for which cryoprotectants like glycerol, xylitol; sugars such as glucose and other precipitants, for example, polyethylene glycol, high salt concentrations are used by adding directly to the crystallization drops or soaking crystals prior to data collection.

6.2 X-ray Crystal Data Collection and the Phase Problem

The collection of X-ray diffraction data starts from a crystal held in a loop at the tip of the cryo-pin being mounted in the path between an X-ray source and an X-ray detector on a magnetic base which is attached to the goniometer head. The diffracted rays coming out of the crystal and originated from the planes of atoms hit the detector (image plate or CCD) where their intensity is accurately measured. The actual X-ray scattering occurs from electrons of molecules. The depiction of the data collection experiment is seen on **Fig. 6-2a**. It is important to collect as complete data as possible, i.e., to record nearly all reflections up to the resolution limit (which is often due to the crystal quality), for which during data collection the crystal is rotated about an axis. Data collection results in a list of images, each representing a wedge of the rotation of the crystal in the beam. A typical wedge, or frame width ranges from $0.2\text{--}1^\circ$. After all reflections have been collected and put on a common scale, using the mathematical technique called Fourier transform the diffraction patterns can be converted into an electron density maps, showing contour lines of electron density. With a good resolution, it is possible to determine atomic positions quite accurately assuming electrons are uniformly distributed around atoms. Each

reflection is characterized by its structure factor. The structure factor, from which one could calculate the electron density distribution in the crystal, is a complex quantity. It has an amplitude and a phase. Only the amplitude, but not the phase can be determined directly from a diffraction experiment. The electron density can be calculated from the structure factors *via* the Fourier transformation using the formula shown below. This task is done nowadays quite fast with the use of computers. The equation, however, contains two unknown quantities, amplitude $|F(hkl)|$ (hkl stands for the reflection indices, or its coordinate in the reciprocal space) and phase of the reflections $\Phi(hkl)$. They must be

$$\rho(x, y, z) = \frac{1}{V_{unitcell}} \sum_{h,k,l} F(h, k, l) \cdot e^{-2\pi i(hx+ky+lz)}$$

$$= \frac{1}{V_{unitcell}} \sum_{h,k,l} |F(h, k, l)| \cdot e^{i\phi} \cdot e^{-2\pi i(hx+ky+lz)}$$

known in order to calculate an electron density map. So, the diffraction experiment directly gives intensities - $|F(hkl)|$, but not phases - $\Phi(hkl)$, which contain the main information about the shape of the molecule. This problem is known as the phase problem in crystallography. The workflow for solving the structure of a molecule by X-ray crystallography is outlined in **Fig. 6-2b**. There are a number of techniques for retrieving the phases from the diffraction data, some of them will be shortly described below [Grune, 2005]:

- 1) Molecular replacement
- 2). Isomorphous replacement

- 3). Anomalous dispersion
- 4). Exploitation of radiation damage

6.2.1 Molecular Replacement

The search of the unit cell with a structure or a fragment of a known structure for the correct orientation and position is undertaken. These coordinates can then be used to calculate first phases for the experimental data. The search includes rotation and

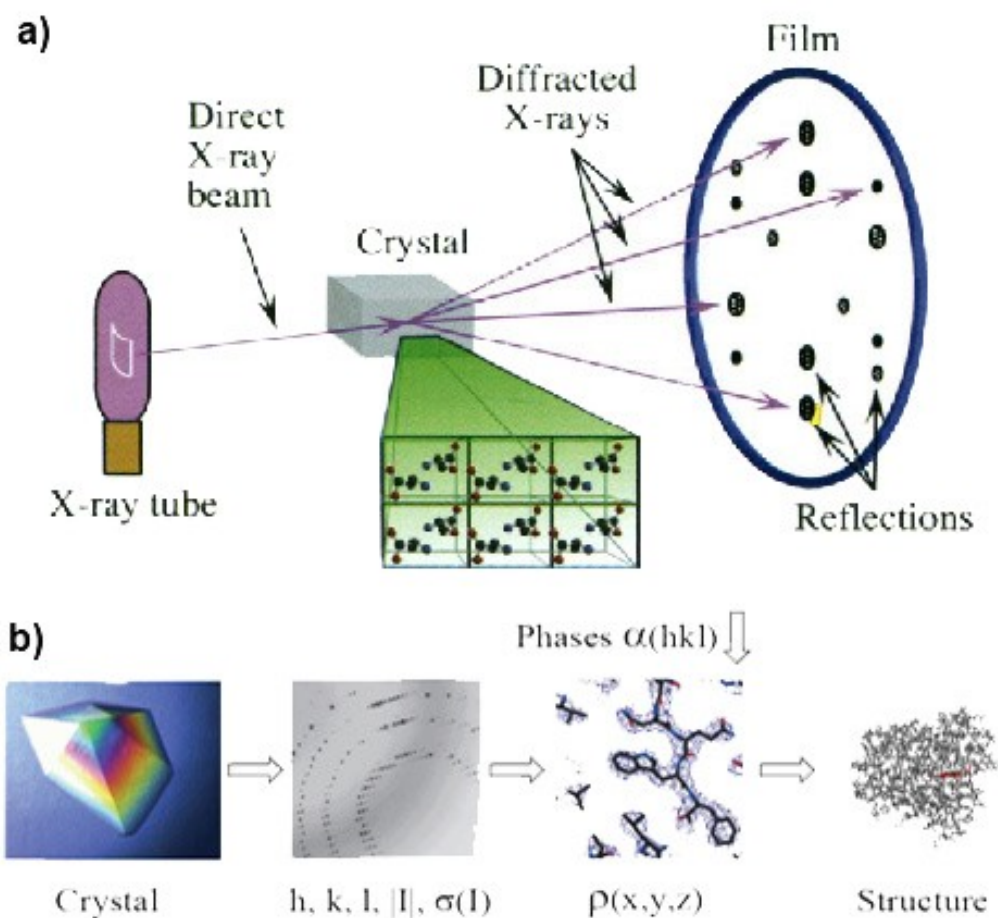


Fig. 6-2 The technique of a single crystal X-ray crystallography. **(a)** Data collection overview. The crystal is hit by the intense X-ray radiation producing diffraction pattern recorded on the detector. [Adapted from Rhodes, 2006]. **(b)** Workflow for solving the structure of a molecule by X-ray crystallography. The first step is to obtain an adequate crystal which is then placed in an intense beam of X-rays, producing the regular pattern

of reflections, the intensity of every reflection is recorded as the crystal is gradually rotated. The data collected with the help of computers result in the electron density map and a model of the arrangement of atoms within the crystal based on it.

translation. **Rotational search.** The Patterson function can be calculated both from the diffraction data and the search model. It does not depend on the position within the unit cell, but only on the orientation. Hence, we can calculate the Patterson for the model in different orientations, compare it with the Patterson of the data, and pick the orientation with the best agreement. **Translational search.** The model is moved through the asymmetric unit keeping the orientation found at the rotational search. At each point, the calculated structure factor amplitudes $|F_c|$ are scored against the observed experimental data $|F_o|$.

The problems associated with the molecular replacement include strong model bias; may sometimes not work even with 100% sequence homology in case of the significant rigid body dislocations (domain movements) [Grune, 2005].

6.2.2 Isomorphous Replacement

Isomorphous replacement from the definition is a method of determining diffraction phases from the differences in intensity between corresponding reflections from two or more isomorphous crystals. Most commonly used in the determination of protein structures, where it is possible to derive isomorphous crystals of native protein and of heavy-atom derivatives (<http://reference.iucr.org/dictionary>). It is based on the introduction of a small molecule into a protein crystal that does not or alter much the structure of the macromolecule. On the other hand, a few heavy metal atoms can contribute detectably to the structure factors and hence introduce changes in the reflection

intensities $|F(hkl)|$. Common heavy metals are Hg, Pb, Au, Pt, Se or U. They can be incorporated by co-crystallisation or by soaking before data collection acquisition. To exploit the method one needs at least two data sets: a native one (no heavy metal) from which the intensities of the native data are obtained $|F_P|$ and a derivative (with heavy metal) to get $|F_{PH}|$. Then one determines the positions of the heavy atoms (x,y,z) (F_H) from the differences $(|F_{PH}| - |F_P|)$, which then allows to calculate phases of a native molecule structure factors Φ_P from $|F_P|$, $|F_{PH}|$ and F_H (**Fig. 6-3**). With a single derivative, the method

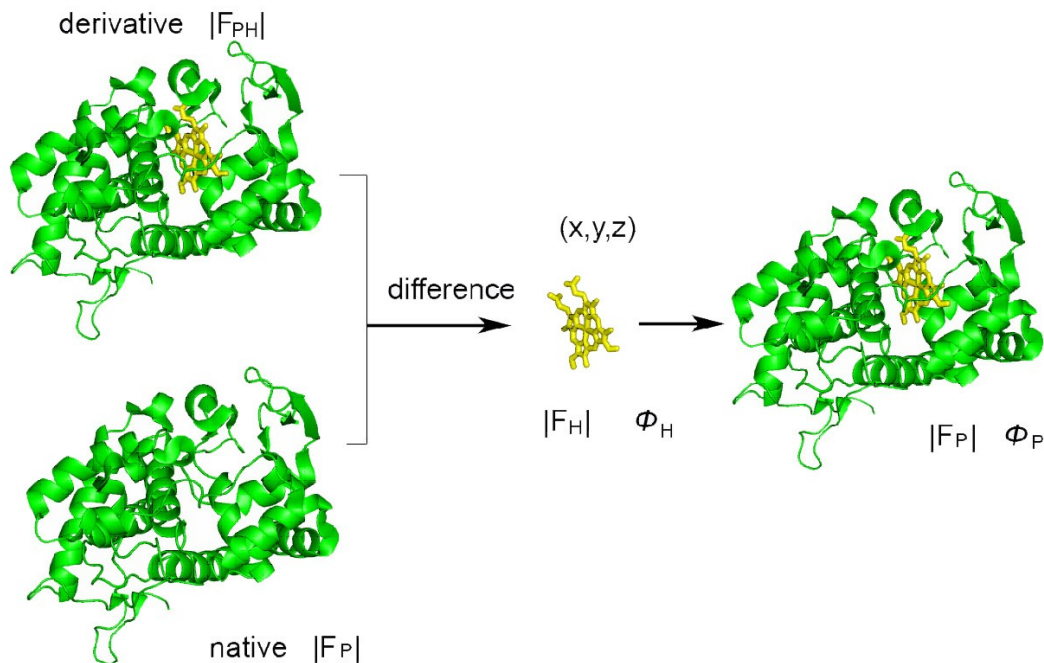


Fig. 6-3 The essence of the isomorphous replacement. This method is based on soaking the crystal of a protein to a heavy atom solution, co-crystallization with the heavy atom. The addition of the heavy atom should not affect the crystal formation or unit cell dimensions in comparison to its native form (isomorphous). From the differences in intensities the structure factor of the heavy alone is calculated which is then used to solve the structure of the protein [Adapted from Grune, 2005].

provides phases for the protein structure up to a twofold ambiguity which could be resolved by preparing the second heavy atom derivative [Grune, 2005; Rhodes, 2006].

6.2.3 Anomalous Dispersion

For a normal diffraction experiment, Friedel's law is valid, which states that the intensities of the reflection (hkl) and (-h-k-l) are equal and that the phases of the underlying structure factor have opposite signs, $\Phi(\text{hkl}) = -\Phi(-\text{h-k-l})$. In a normal diffraction experiment an X-ray wave impinges on an atom and causes the electrons to oscillate in the same frequency as the incoming X-ray wave. However, in anomalous scattering experiment, when the energy of the incident X-ray wave is close to an element absorption edge, in addition to being scattered elastically, the incident X-ray wave is absorbed and an element ejects a core electron which then falls back to a lower shell emitting X-ray radiation. To maximize the anomalous scattering the wavelength of an X-ray must be tuned to be absorbed by a particular atom. Determining the absorption edge is normally done before data collection acquisition by a fluorescence scan (scattering of X-rays at right angle to the incident beam). Several (MAD) or only one wavelength at the absorption peak (SAD) experiment data are collected to exploit the difference in Friedel's pairs and dispersive differences (intensities vary slightly at different wavelengths) to find phases.

Some other techniques for obtaining the phases have been successfully exploited such as RIP and Sulphur-SAD. RIP makes use of the fact that radiation forms radicals, which damage the molecule, and apart from random destruction, carboxyl-groups are removed and disulphides are destroyed. To use RIP, a normal data set is collected ("native"), then the crystal is exposed to a high dose of X-rays, to damage the crystal, then a second set is collected, which would anticipate that the chemical changes have been caused by the radiation damage and the data are treated as derivative. Sulphur-SAD uses the very weak

signal of native S, P for nucleic acid structures or any other light atoms present in organic molecules [Grune, 2005; Rhodes, 2006].

7. PRACTICAL METHODS

7.1 Expression and Purification of HsdR

7.1.1 Expression and Purification of Native Recombinant HsdR

R.EcoR124I was expressed in BL21(DE3) *E. coli* (Novagen) from the plasmid pTRC124 [Janscak and Bickle, 2000]. The overnight culture of cells was diluted 1:100 in fresh 1 liter LB medium in a 3 liter conical flask supplemented with ampicillin to a final concentration of 100 $\mu\text{g ml}^{-1}$ and grown with shaking at 310 K until OD_{600} was about 0.4-0.5 following the induction with isopropyl β -D-thiogalactopyranoside (IPTG) to a final concentration of 0.8 mM and the culture was incubated additional three hours at 290 K. Cells were harvested (4500 rpm 20 min) and resuspended in 60 ml of buffer A (20 mM potassium phosphate pH 7.5) and broken down by sonication on ice for 2 x 30 s bursts with 30 s cooling in between. The cell lysate was clarified by ultracentrifugation (25000 rpm for 1h at 277 K). The supernatant was directly applied onto a 20 ml DEAE Hi-Trap Sepharose Fast Flow column (Amersham) preequilibrated in buffer A. Chromatographic process was performed using FPLC machine (Pharmacia). Bound proteins were eluted with a linear gradient of KCl (0.01 M-0.5 M, 600 ml). Fractions containing recombinant protein were pooled and concentrated to 16 mg ml^{-1} using 50 kDa cut-off spin

concentrators (Millipore). Protein concentration was calculated from absorbance at 280 nm using molar extinction coefficient derived from amino acid sequence (98225/M/cm) [Gill and von Hippel, 1989] using Protein Identification and Analysis Tools on the ExPASy Server (<http://www.expasy.org/>). Prior to crystallization the enzyme solution was supplemented with ATP to final concentration of 5 mM. Protein was stored at 250 K.

7.1.2 Expression and Purification of the SeMet-labeled Recombinant HsdR

The plasmid pTrcR124 containing the *hsdR* gene under the control of the Ptrc promoter was used to over-produce the HsdR subunit of EcoR124I (R.EcoR124I). Selenomethionine-derivatized R.EcoE124I protein was produced by transforming the pTrc124 plasmid into the auxotrophic strain *E. coli* B834 (DE3) (Novagen). The overnight cell culture was grown in 10 ml minimal M9 media containing L-selenomethionine (50 $\mu\text{g ml}^{-1}$; Sigma) and ampicillin (100 $\mu\text{g ml}^{-1}$) at 310 K, 180 rev. min^{-1} . Overnight culture then was diluted 1:100 in fresh 1 l M9 minimal media of the same content and grown with shaking at 310 K, 180 rev. min^{-1} until OD₆₀₀ (optical density at 600 nm) was about 0.35 – 0.4, whereupon the solution was cooled to 290 K and protein overexpression was induced by the addition of 0.8 mM isopropyl β -D-thiogalactopyranoside (IPTG) (final concentration). After 24 h, cells were harvested (4500 rpm, 20 min, 277 K) and resuspended in 60 ml of buffer A (20 mM potassium phosphate pH 7.5) and were disrupted by passing the suspension twice through a French pressure chamber (THERMO electron corporation, 40K cell, USA) with the internal cell pressure 103 MPa. The cell lysate was clarified by ultracentrifugation (25000 rpm for 1h at 277 K, Beckman). The supernatant was directly applied (1 ml min^{-1}) onto a 20 ml

DEAE Sepharose Fast Flow column (Amersham Bioscience) preequilibrated with buffer A. The column was attached to the FPLC machine (Pharmacia) and washed with buffer A (4 ml min^{-1}) till no protein peaks were detected and subsequently washed with buffer A supplemented with 5 mM KCl . Bound proteins were eluted with a linear gradient of KCl (0.01 M - 0.5 M , 4 ml min^{-1} , 600 ml). Fractions were analyzed by 10% SDS-PAGE (Bio-Rad). Recombinant HsdR was eluted from the DEAE Sepharose column as a single peak, monitored using dual path UV monitor (Pharmacia). Some amount of peak fractions containing recombinant protein was pooled and concentrated to 15 mg ml^{-1} using 50 kDa cut-off spin concentrators (Millipore). The rest was dialyzed against the buffer A and applied to the MonoQ column, with the following procedure of purification as on the DEAE Sepharose described previously. On the next step the gel filtration chromatography on the Sephadex G75 was performed and prior to crystallization the enzyme solution was supplemented with ATP to final concentration of 5 mM with the protein being stored at 250 K .

7.3 Crystallization of the Native and Labeled HsdR

In case of the **native HsdR**, the 15 mg ml^{-1} protein solution consisting of the recombinant HsdR in 20 mM phosphate buffer pH 7.5 , 0.1 M KCl and 5 mM ATP was used for the crystallization experiments. Crystallization trials were performed in Hampton Research Linbro plates (Hampton Research, California, USA) at both room temperature and at 277 K using a sitting-drop vapor-diffusion method. Commercial crystal screen kits from Hampton Research (CA, USA) and Molecular Dimensions Limited (Suffolk, UK) were used for initial screening of crystallization conditions.

Crystals of the **SeMet-labelled** R.EcoR124I were grown using the sitting-drop vapour-diffusion technique in the presence of ATP by mixing 1 μ l protein solution (15 mg ml⁻¹, 20 mM potassium phosphate pH 7.5, 100 mM NaCl, 5 mM ATP) with 2 μ l reservoir solution (500 μ l), composed of 0.2 M Li₂SO₄, 8 % PEG 20 000, 8 % PEG 550 MME, 1.5 mM DTT.

7.4 X-ray Data Collection

7.4.1 X-ray Data Collection on Native HsdR Crystals

For data collection native crystals were mounted in a nylon loop (Hampton Research) and transferred for a few seconds to 10 μ l of a cryosolution containing 25% glycerol (in water). The crystals were then flash-frozen in a cold nitrogen stream at 100 K for diffraction experiment. Data were collected at wavelength 1.0 \AA using synchrotron radiation at the X12 beamline of the EMBL Hamburg Outstation. A total of 720 images were recorded with an oscillation angle of 0.5° and an exposure time of 60 s per image using a 225 mm MAR Mosaic CCD detector at the beamline. Crystal to detector distance was set to 150 mm. The intensity data were processed and scaled using the HKL package [Otwinowski and Minor, 1997].

7.4.2 X-ray Data Collection on Labeled HsdR Crystals

Data collection of the labeled crystals were carried out at 100K at the energy-tunable beamline X12 at the EMBL Hamburg Outstation. A total of 460 frames of data were collected on a MAR Mosaic 225mm CCD area detector at the experimentally determined selenium absorption-edge peak energy (12.671 keV, 0.9784 \AA) with an oscillation angle

of 0.5°, exposure of 90 s with a crystal-to-detector distance of 250 mm. Images were processed using DENZO and SCALEPACK [Otwinowski and Minor, 1997].

7.5 Structure Determination

The structure was solved using SeMet Single wavelength Anomalous Dispersion (SAD) method. Substructure determination, phase calculation, solvent flattening, and partial model building were performed automatically by using SHELXD [Schneider and Sheldrick, 2002], SHARP [La Fortelle and Bricogne, 1997], DM [Cowtan, 1994], and ARP/wARP [Perrakis, 1999], respectively, within the software pipeline AutoRickshaw [Panjikar et al., 2005]. Further manual model building was performed using Coot [Emsley and Cowtan, 2004]. Initial rounds of refinement were carried out applying restrained noncrystallographic symmetry using the CNS program package [Brünger et al., 1998]. Libraries and CNS input files for the inclusion and energy minimization of ATP were generated using PRODRG [Schuettelkopf van Aalten, 2004] and the HIC-Up server [Kleywegt, 2007] and used in CNS refinement. Finally, restrained refinement by maximum likelihood in REFMAC5 [Murshudov et al., 1999] was combined with TLS [Winn et al., 2001] refinement. TLS and restrained refinement consisted of 10 cycles with atomic residual isotropic *B* factors set at 20 Å². TLS groups were calculated by analyzing the spatial distribution of individual atomic thermal parameters using TLSMD web server [Painter et al., 2006]. The structure was analyzed using PROCHECK [Laskowski et al., 1993].

Figures were prepared with PyMOL (<http://pymol.sourceforge.net/>). Secondary Structure Matching was performed using SSM tool at EBI (Protein structure comparison service SSM at European Bioinformatics Institute <http://www.ebi.ac.uk/msd-srv/ssm>, authored

by E. Krissinel and K. Henrick). Sequence-homology analysis was performed with ClustalW [Higgins et al., 1994].

7.6 Predictions and Modeling

Sequence-homology analysis was performed with ClustalW [Higgins et al., 1994]. Three-dimensional alignments were performed in the swiss pdb viewer [Guex and Peitsch, 1997] and in Yasara [Krieger et al., 2004] using Sheba [Jung and Lee, 2000]. For secondary structure predictions of the complete helical domain the following algorithms were used: JPRED, Jnet, jhmm, jpssm [Cuff and Barton, 1999], PHD and PROFsec [Rost and Sander, 1994], PSI_PRED [Jones, 1999], NNpredict [McClelland and Rumelhart, 1988; Kneller et al., 1990]. Actual model building by joining the three-dimensional aligned backbones of the HsdR crystal structure and the M-helical domain to guide the fold of the unresolved part of the helical domain was performed in Yasara. During model refinement and replacement of the M-helical sidechains with HsdR sidechains the structure was continuously minimized with the Amber99 force field [Wang et al., 2000].

III. RESULTS

8. Expression and Purification of HsdR

8.1 Expression and Purification of the Native Recombinant HsdR

Recombinant R.EcoR124I was expressed in BL21(DE3) *E. coli* in 1 liter LB medium. The fact that recombinant HsdR was expressed in a big quantities could be visually evaluated comparing the appearance of the cell before and after expression (**Fig. 8-1**). Ion-exchange DEAE Hi-Trap Sepharose chromatography was then used. Fractions from the column were analyzed by 10 % SDS-PAGE [Maniatis et al., 1982] and showed the purity of the peak fraction was more then 95% (**Fig. 8-2a**). Recombinant HsdR was eluted from the DEAE Sepharose column as a single peak, monitored using dual path

Fig.8-1 The comparison of the BL21(DE3) *E. coli* cells before (**a**) and after (**b**) the HsdR expression.

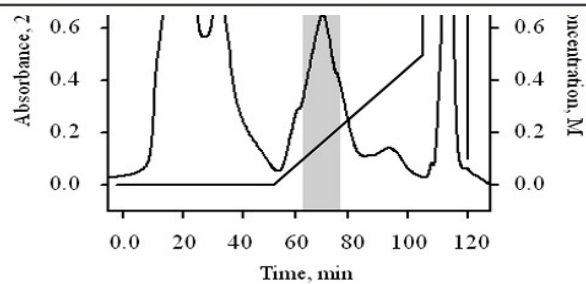
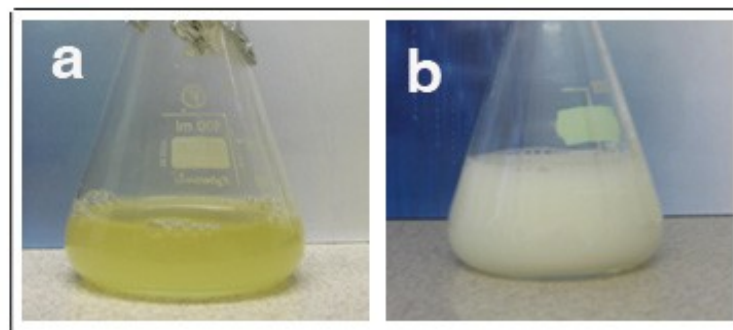
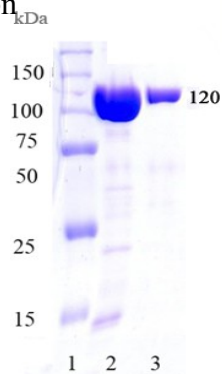


Fig. 8-2 (a) SDS-PAGE (12%) showing the purification of the R.EcoR124I. Lane 1, molecular mass markers; lane 2, 3, purified R.EcoR124I. **(b)** Elution profile from purification of HsdR subunit of the EcoR124I endonuclease on DEAE Sepharose Fast Flow. System: FPLC (Pharmacia). Shaded area represents peak fraction containing R.EcoR124I.

UV monitor (Pharmacia) (**Fig. 8-2b**). Fractions containing recombinant protein were pooled and concentrated to 16 mg ml⁻¹, supplemented with ATP to final concentration of 5 mM and stored at 250 K

8.2 Expression and Purification of the SeMet-labeled Recombinant HsdR

Selenomethionine-derivatized HsdR of EcoE124I protein was produced by transforming the pTRC124 plasmid into the auxotrophic strain *E. coli* B834 following the overnight cell culture grown in minimal M9 media containing L-selenomethionine and induction of the overexpression by the addition of isopropyl β -D-thiogalactopyranoside (IPTG). Disrupted by the French pressure chamber cell lysate was clarified and directly applied onto the DEAE Sepharose Fast Flow column. Recombinant HsdR was eluted from the DEAE Sepharose column as a single peak, monitored using dual path UV monitor. Eluted fractions were analyzed by 10 % SDS-PAGE which showed that the purity of the peak fraction was more than 95% (**Fig. 8-3**). Some amount of peak fractions containing recombinant protein was pooled and concentrated to 15 mg ml⁻¹ while the rest was dialyzed and applied to the MonoQ column, with the following procedure of purification

as on the DEAE Sepharose described previously. On the next step the gel filtration chromatography on the Sephadex G75 was performed. On all steps the eluted fractions were analyzed by 10 % SDS-PAGE (**Fig. 8-3**) and prior to crystallization the enzyme solution was supplemented with ATP to final concentration of 5 mM with the protein being stored at 250 K.

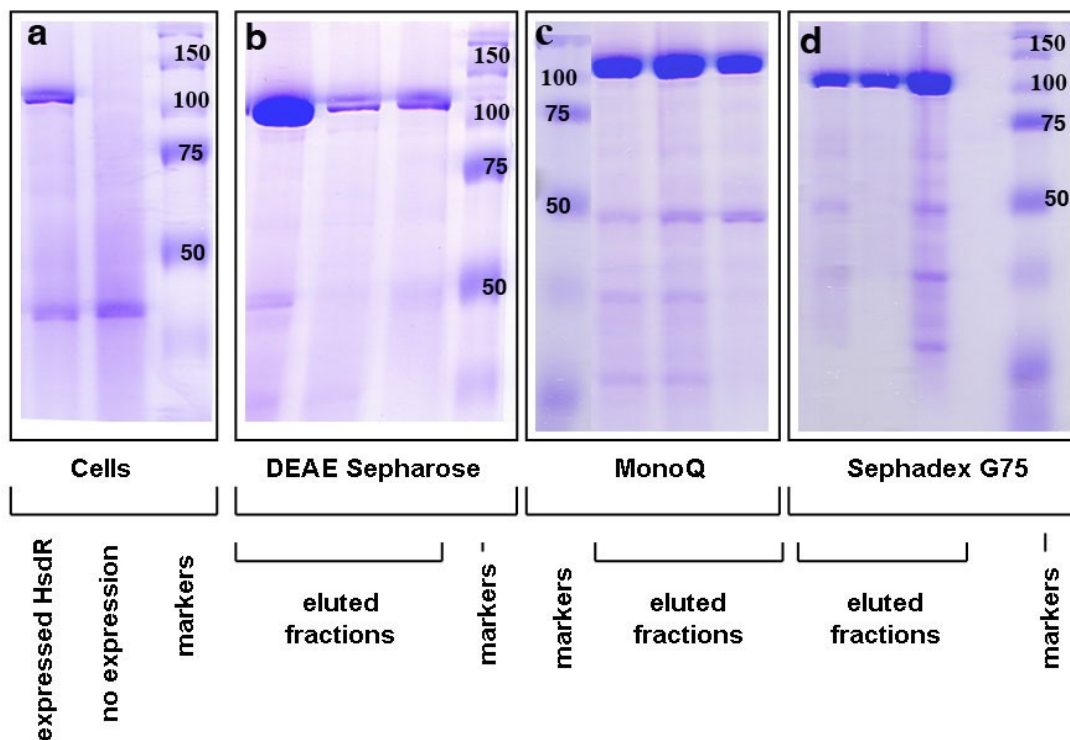


Fig. 8-3 Expression and purification of the SeMet HsdR subunit of EcoR124I enzyme. Presence of the protein after expression in *E. coli* B834 (DE3) cells (**a**) and its purity at a different purification steps: (**b**) DEAE Sepharose, (**c**) MonoQ column ion exchange chromatography and (**d**) after gel filtration were analyzed by SDS-PAGE and Coomassie staining. Molecular weight markers (in kDa) shown for every panel separately.

9. Crystallization of the Native and Labeled HsdR

Crystals of the recombinant native HsdR protein were grown from mixture of polyethylene glycol solutions, yielding different crystal forms (**Fig. 9-1a-j**). The best crystals were obtained from precipitant (500 μ l reservoir solution) containing 8 % PEG 20 000, 8 % PEG 550 MME and 0.2 M Li_2SO_4 at 277 K. The crystallization drops consist of 2- μ l protein solution (protein in 20 mM potassium phosphate pH 7.5, 100 mM KCl, 5 mM ATP in water), 2- μ l precipitant and 0.5 μ l of 0.1 M CoCl_2 as an additive. Crystals appeared within a few days and grew to maximum dimensions of about 0.5 x 0.4 x 0.1 mm (**Fig. 9-1a**).

Apoprotein (protein without ATP) was also successfully crystallized using condition No. 23 (0.2 M KSCN, 8% PEG 20 000, 8% PEG 550 MME) of Molecular Dimensions Clear Strategy Screen 1 (Molecular Dimensions Limited). The initial crystals with dimension of about 0.06 x 0.02 x 0.02 mm (**Fig. 9-1b**) were not suitable for diffraction experiments. Further improvement of crystallization conditions did not lead to satisfactory results until ATP complex was used for experiments. Crystallization of protein complex with ATP led to the rapid improvement of crystals size and shape. Crystals grown from precipitant consisted of 8 % PEG 20 000, 8 % PEG 550 MME and 0.2 M Li₂SO₄, (condition No. 20, CSS-I, MD1-14 from Molecular Dimensions Limited) at 277 K (**Fig. 9-1c**). Plate crystals appeared together with numerous intergrown crystals in a single drop. The Hampton Research Additive Screen kit was used for screening different ions to improve the quality of crystals (Additive Screen HT, HR2-138 of Hampton Research, California, USA). Crystals grew in the presence of different ions but improvement of diffraction data quality was achieved only when 0.1 M. CoCl₂ (Additive Screen HT, HR2-138, condition 4) was added into the drop to the final concentration of 11 mM.

Crystallization conditions were optimized for **SeMet-labelled HsdR** protein based on those for the native HsdR [Lapkouski et al., 2006]. Crystals of the SeMet-labelled R.EcoR124I were grown using the sitting-drop vapour-diffusion technique in the presence of ATP. Boulder-shaped crystals grew in one week in precipitant composed of 0.2 M Li₂SO₄, 8 % PEG 20 000, 8 % PEG 550 MME, 1.5 mM DTT (**Fig. 9-1k**). The best crystals were obtained from the protein fractions coming from the DEAE Sepharose purification step.

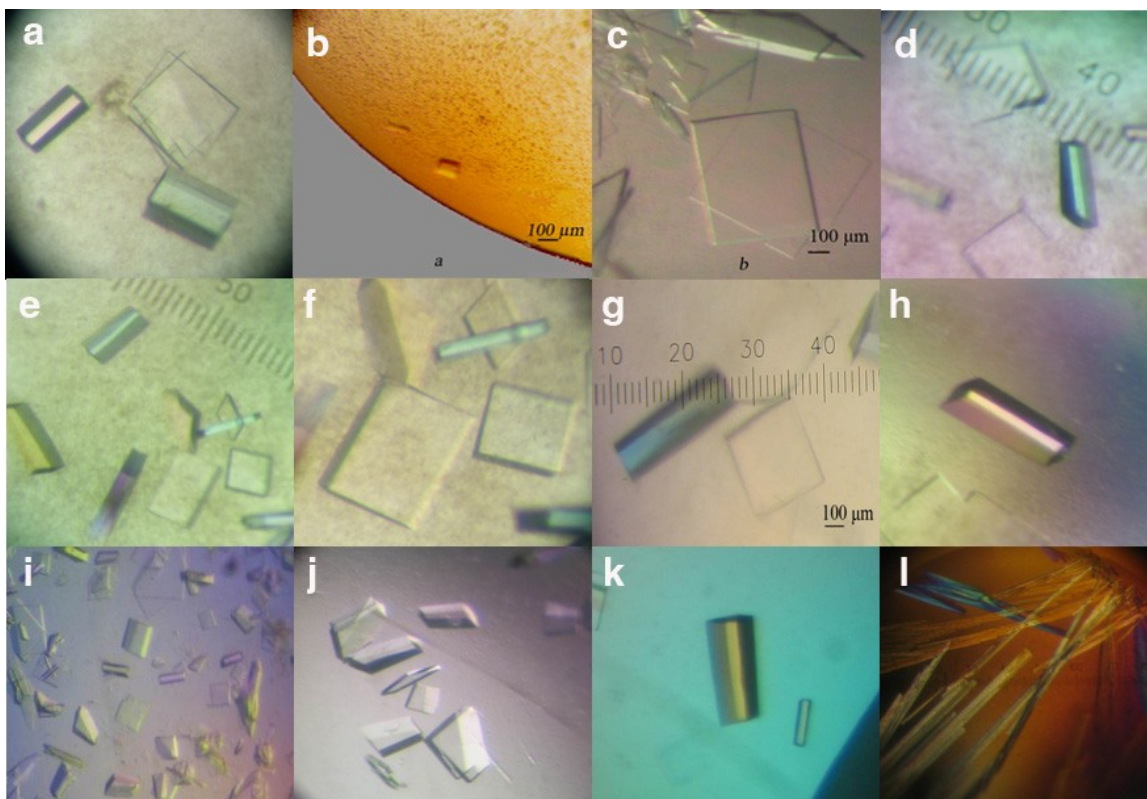


Fig. 9-1 Different crystal forms of the R.EcoR124I protein appeared during the screening of crystallization conditions; **(a)** containing 8 % PEG 20 000, 8 % PEG 550 MME and 0.2 M Li_2SO_4 at 277 K. with a maximum dimensions of about 0.5 x 0.4 x 0.1 mm **(b)** Crystal of the apoprotein (protein without ATP) in 0.2 M KSCN, 8% PEG 20 000, 8% PEG 550 MME, 298 K, **(c)** Numerous plates growing within 30 min from the protein with 5 mM ATP. Precipitant consist of 8 % PEG 20 000, 8 % PEG 550 MME, 0.2 mM Li_2SO_4 at 298 K, **(d-j)** Crystals of a protein with ATP appearing together in a single drop in the presence of 11 mM CoCl_2 at 277 K **(k)** Crystals of the SeMet-labelled R.EcoR124I with ATP, grown in precipitant, containing 0.2 M Li_2SO_4 , 8 % PEG 20 000, 8 % PEG 550 MME, 1.5 mM DTT. **(l)** Salt crystals of ammonium sulfate $(\text{NH}_4)_2\text{SO}_4$.

10. Preliminary X- ray Diffraction Analysis of the Native and Labeled HsdR

As noted above, the initial **native crystals** with dimension of about 0.06 x 0.02 x 0.02 mm (**Fig. 9-1b**) were not suitable for diffraction experiments. Thin plate crystals produced as aggregates and intergrown forms in a single drop (**Fig. 9-1c**) diffracted to a maximum resolution of 2.8 Å but due to high mosaicity and blurred spots it was not

possible to process the diffraction data (**Fig. 10-1a**). Native crystals produced in the presence of 0.1 M CoCl₂ (**Fig. 2d-j**) yielded a reasonable diffraction quality. Native diffraction data (**Fig. 10-1b**) were collected at wavelength 1.0 Å to resolution 2.6 Å. Crystals belong to the monoclinic space group, with unit-cell parameters a = 85.75 Å, b = 124.71 Å, c = 128.37 Å, β = 108.14° with possibly two or three molecules in the asymmetric unit. The statistics of the crystallographic data are summarized in **Table 2**. Protein sequence search against the protein data bank did not reveal any structural homology, therefore, molecular replacement method was not attempted and selenomethionine-derived protein was produced to solve the structure using single/multiple anomalous diffraction (SAD) experiment.

For the **SeMet-labelled** HsdR, PEGs composition of the drops was not sufficient for an optimal cryoprotection, therefore crystals of SeMet R.EcoR124I were cryoprotected by soaking in 18 % (v/v) 2-methyl-2,4-pentanediol (MPD) in water. Crystals were mounted in a nylon loop and flash cooled for data collection, giving a data set to 2.6 Å resolution (**Fig. 10-1c**) with a redundancy of 9.4 and an overall R_{merge} of 13 %. The crystal belonged to the space group *P2₁2₁2₁*, with unit-cell parameters a = 123.91, b = 129.92, c = 161.01 Å. Based on the molecular weight and the space group, it was assumed that the crystal contained two protein molecules per asymmetric unit, giving a Matthews coefficient [Matthews, 1968] (*V_M*) value of 3.1 Å³ Da⁻¹ and a solvent content of 54% (**Table 2**).

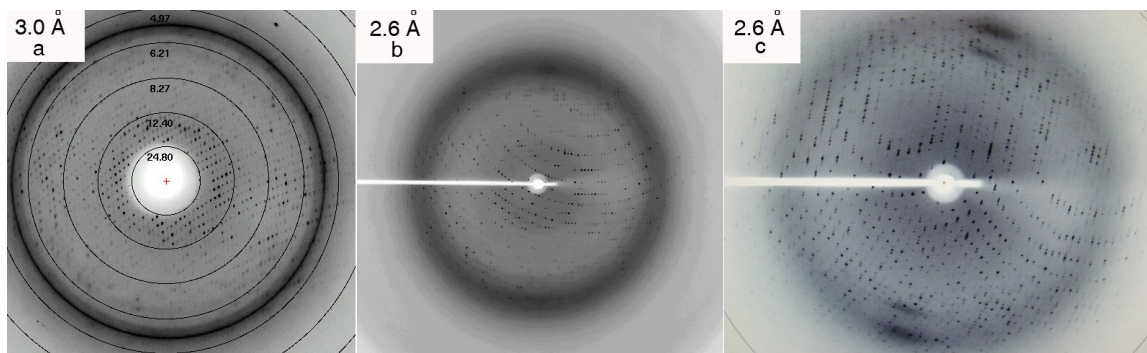


Fig. 10-1 Diffraction images of the native (a,b) and SeMet-labelled (c) HsdR enzyme. Crystals diffracted to the resolution of 3.0 Å, 2.6 Å and 2.6 Å respectively.

Table 1

Data collection statistic for the R.EcoR124I native and SeMet derivative crystals. Values in parentheses correspond to the highest resolution shell.

Data set	Native HsdR	SeMet-HsdR
Beamline	X12	X12
λ (Å)	1.0	0.9784
Resolution range (Å)	20 – 2.6 (2.64-2.60)	20.0–2.60 (2.64-2.60)
Unit-cell parameters	a = 85.75 Å, b = 124.71 Å, c = 128.37 Å, $\alpha = 108.14^\circ$	a = 123.91 Å, b = 129.92 Å, c = 161.01 Å, $\alpha, \beta, \gamma = 90^\circ$
$\Delta\phi$ (°)	0.5	0.5
Space group	P2/P2 ₁	P2 ₁ 2 ₁ 2 ₁
Redundancy	7.3 (6.7)	9.4 (9.1)
Measured reflections	582265	758388
Unique reflections	79374	80316
Completeness (%)	99.3 (96.5)	100 (100)
Rmerge ^a (%)	10.9 (55.0)	13 (49.7)
I/σ(I)	18.2 (2.9)	17.9 (5.8)
Mosaicity (°)	0.99	0.6

^aRmerge = $\frac{\sum_{hkl} \sum_i |I_i - \langle I \rangle|}{\sum \langle I \rangle}$, where I_i is the intensity of the i th measurement of reflection hkl and $\langle I \rangle$ is the average intensity of a reflection.

11. STRUCTURE OF THE MOTOR SUBUNIT OF ECOR124I ENZYME

11.1 Structure Determination and Quality

SAD data gave an interpretable electron density map, which was successfully used for a model building. An anomalous difference Fourier peak near the ATP was interpreted as Mg²⁺ on the basis of its coordination number and geometry. Magnesium was not used in

the preparation of the protein solution. The enzyme appeared to have picked up them during the expression or purification procedures or comes as an impurity from chemicals. The final R_{cryst} and R_{free} of the model were 22.1% and 26.5% respectively. In the Ramachandran plot, 85.6 %, 13.1 % and 1.1 % of the residues (1285, 197 and 17 amino-acid residues, respectively) are in the most favoured regions, additional allowed regions and generously allowed regions, respectively. 3 residues are in the disallowed region. The final refined model consists of 1668 amino acid residues, two ATP molecules in the active site, 2 magnesium ions and 514 water molecules. Two molecules in asymmetric unit with residues 1-12, 142-147, 182-189, 583-590, 857-869, 886-1038 of the molecule A and 1-12, 142-147, 182-190, 585-590, 863-869, 893-1038 of the molecule B are disordered.

11.2 Subunit Architecture: A Planar Array of Functional Domains

HsdR is organized into four approximately globular structural domains in nearly square-planar arrangement (**Figure 11-1**). Similarity searches (using Secondary Structure Matching at <http://www.ebi.ac.uk>) reveal no significant matches to HsdR itself, but each domain resembles proteins with functions similar to known properties of the motor subunit. The N-terminal domain (residues 13-260) resembles endonuclease domains conserved among other RMs; the second (261-461) and third (470-731) domains resemble typical recA-like helicase domains 1 and 2.

The near-planar arrangement of globular domains creates prominent grooves between each domain pair. At the top of the subunit in the view of **Figure 11-1** the two helicase-like domains form a canonical helicase cleft very similar to those found in helicases or translocases of known structure [Dürr et al., 2005]; in the HsdR structure double-stranded

B-form DNA can be accommodated in this cleft without steric clash as in co-crystals of other helicases with bound nucleic acid substrates. At the bottom of the subunit the

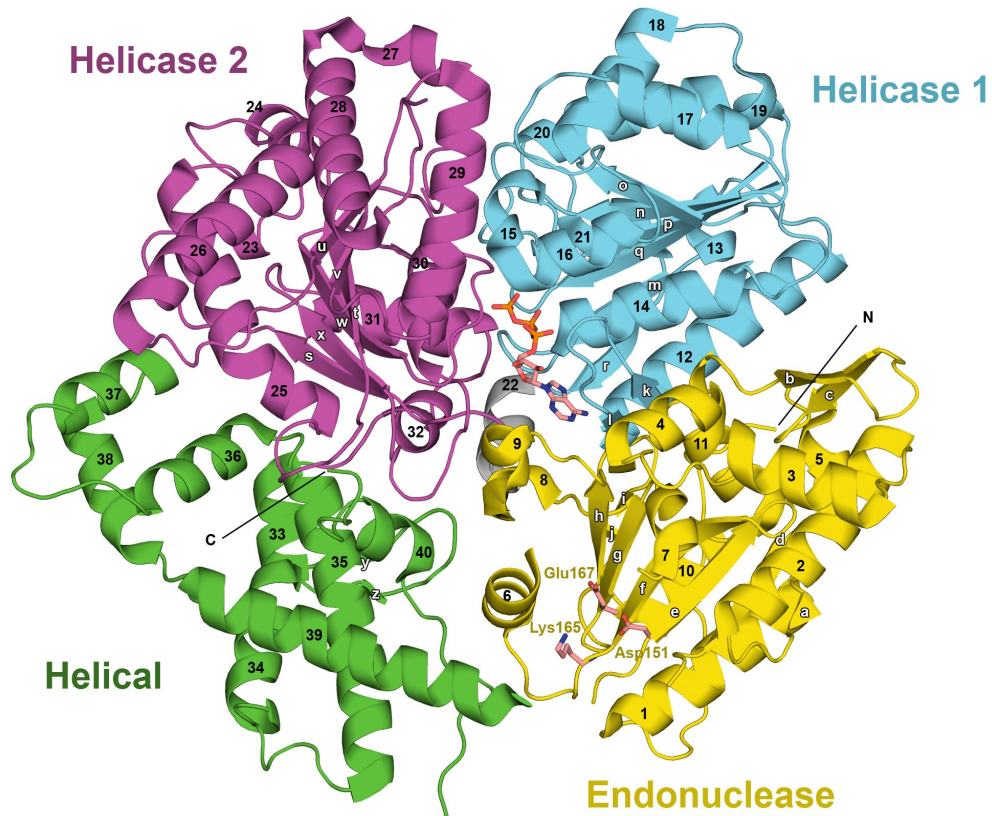


Fig. 11-1 Structure of the motor subunit. Planar domain assembly. 'Front' view indicating domain names and colors. Helices are numbered 1 to 40 (black) and strands are lettered a to z (white) sequentially from N- to C-terminus (labeled). Dashed lines represent short regions of disorder; unresolved residues 892 to 1038 are not shown. Endonuclease active-site residues and ATP are shown as skeletal models in atomic colors with gold carbon, Mg²⁺ as a black sphere.

endonuclease domain faces the helical domain across a gap that appears to be due at least in part to the ~150 residues that are disordered at the C-terminus. Both endonuclease and helical domains present small domains near the 'waist' of the subunit that form twin knobs protruding slightly from the plane, respectively up and down in the view of **Figure 11-1**. All four domains are quite similar in size, and in **Figure 11-1** each is oriented

approximately with its smallest dimension in the plane of the page. The thickness of the subunit ranges from only ~35 to ~60 Å, whereas its height (~90 Å) and width (~85 Å)

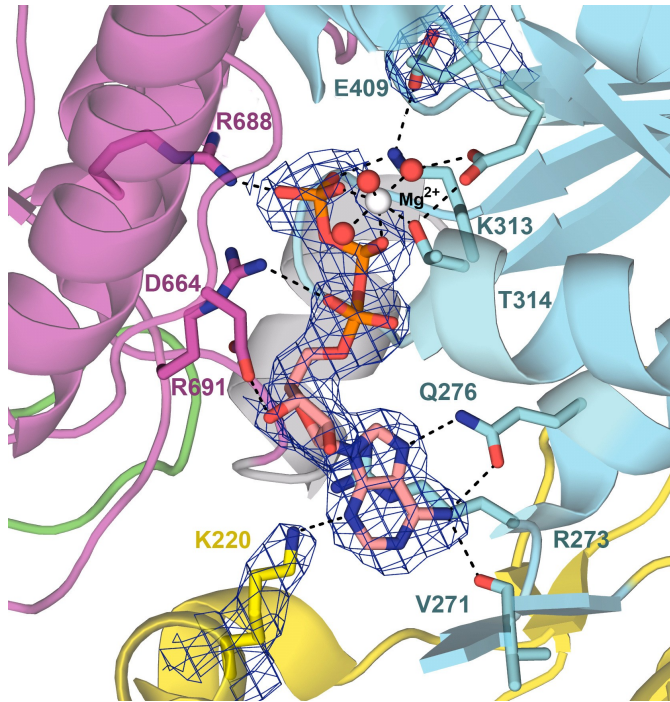


Fig. 11-2 ATP binding site details. Zoom-in view of panel a except with Mg^{2+} as a white sphere and with ATP electron density added as blue mesh. Dashed lines represent atomic contacts discussed in the text with residues labeled in one-letter code shown in atomic colors with carbon in domain color. Side chains of the several important in catalysis residues are shown with the corresponding electron density map.

are nearly equal. A counterintuitive consequence of a planar arrangement of approximately equal-sized globular domains is that each domain buries more surface area with its neighbors than it would if more closely packed, as, e.g., in a tetragonal arrangement, but the subunit as a whole is less compact. Although three of the four domains have mixed α/β folds, nearly all interdomain interfaces are made by helical segments that surround the β strands. Interdomain contacts are also mediated by ATP, which is deeply buried between helicase domains 1 and 2 except for its exposed N3 edge facing the endonuclease domain (**Fig. 11-2**).

11.2.1 Conservation and Divergence in the Endonuclease Domain

The endonuclease domain of HsdR presents an $\alpha/\beta/\alpha$ architecture with a central, curved, six-stranded sheet flanked on both sides by helix bundles. This fold shares a common core (here strands d through h and helices 1, 10, and 11) with the endonuclease subunits

of most type II REases of known structure (**Fig. 11-3**) [Niv et al., 2007]. As in the type II restriction enzymes, the spatial order of the core strands in the sheet is identical to their

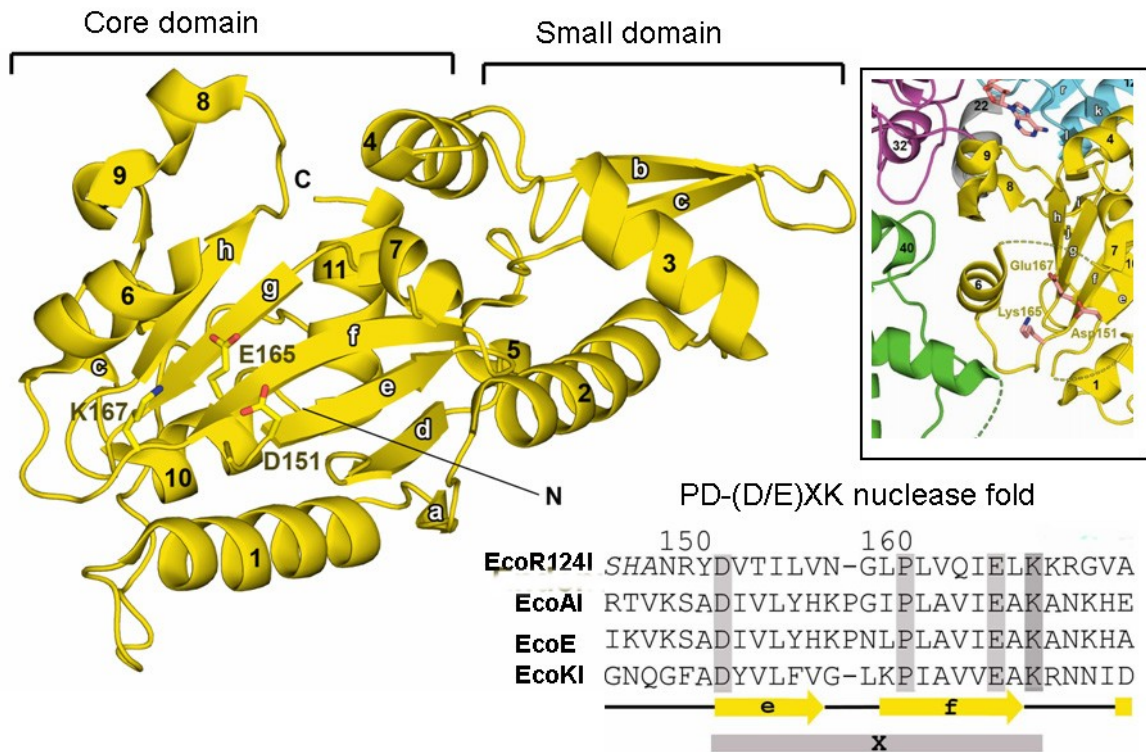


Fig. 11-3 The structural features of the Endonuclease domain. The domain could be divided into two parts: 1) the core domain, where the PD-(D/E)XK motif, a representative motif of type II restriction enzymes, is located; and 2) the smaller domain, having no spatial similarity to other endonucleases, function of which remains unknown. In the insert panel the relative position of the HsdR cleavage site in regard to the helicase domains (red and cyan) and Helical domain (green) represented.

linear order in the primary structure. In addition to the core elements, HsdR helices 6 and 9 correspond to a pair of helices found on the convex face of the sheet in most endonucleases of known structure. Residues previously proposed to be involved in catalysing DNA cleavage cluster near the center of the domain array on secondary structure elements of the $\alpha/\beta/\alpha$ core, where they face the helical domain across the disordered gap (**Fig. 11-3**). Asp151, Glu165, and Lys167 are the residues identified by similarity to the catalytic sequence motif PD-(D/E)XK conserved among type I (**Fig. 11-**

3), II, and III RMs [Titheradge et al., 1996; Davies et al., 1999]. Disordered segments ~142-147 and ~182-190 also face the active site region. The latter segment presents at the end of helix 6 residues Gln179 and Tyr183 that are proposed to play a role in catalysis [Sisakova et al., 2008].

The endonuclease domain can make a novel contact to the ATP bound at the helicase domains. The exposed N3 edge of the adenine ring faces the amino protons of Lys220 on $\alpha 8$ (**Figure 11-2**) within hydrogen-bonding distance (3.1 Å). Coordination of this ring position is not observed in other helicase-like proteins of known structure, according to a survey of the PDB conducted in June, 2008 (not shown).

11.2.2 Functionally Integrated Helicase Domains

HsdR helicase domains 1 and 2 together resemble a number of proteins or domains with parallel, six-stranded, RecA-like α/β folds typical of so-called DEAD-box helicases [recently reviewed by Singleton et al., 2007]. Structural superpositioning identifies the most similar relatives of the HsdR helicase domain pair as human eIF4AIII (PDB ID 2j0s, as a part of the Exon Junction Complex, EJC) [Bono et al., 2006], a subunit of the RNA chaperone exon-junction complex; Vasa (PDB ID 2db3) [Sengoku et al., 2006], a *Drosophila* RNA-unwinding protein; and Rad54, a ubiquitous recombination translocase (PDB ID 1z63 is from *Sulfolobus solfataricus*) [Durr et al., 2005]; each of these three superimposes ~300 residues on HsdR with ~3 Å rmsd. Although pairs of helicase domains invariably occur together and have approximate two-fold symmetry that mimics a homodimer, the domains are pseudosymmetric due to their being confined to the same chain, and bind only one ATP between them. Among available helicase crystal structures the angles and distances between the two helicase domains vary considerably, showing

no unique correlation with ligand occupancy by ATP, ADP, or their analogs, or nucleic acid substrates. The ATP-bound structures of UvrB [Theis et al., 1999], PcrA [Velankar et al., 1999], and the RecQ catalytic core [Bernstein et al., 2003] have been described as being in active DNA-binding conformations. The ATP-bound HsdR helicase domains are also apparently in an active state for DNA binding, judging by comparison with the RNA-bound states of Vasa and exon-junction helicases in the absence of ATP. Thus, an active state for DNA transport apparently can be captured in crystals without simultaneous

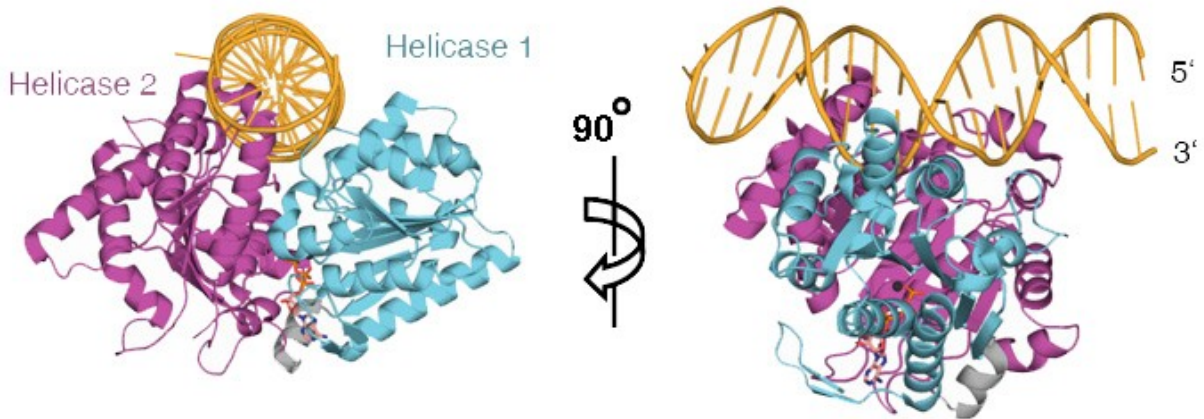


Fig. 11-4 Model of the complex of HsdR and dsDNA. The dsDNA could be accommodated between two helicase domain without considerable sterical hindrances.

binding of both ATP and nucleic acid substrates. As in the DNA-bound structures of RecG (PDB 1gm5) [Singleton et al., 2001], PcrA (PDB 3pjr), NS3 (PDB 1a1v) [Kim et al., 1998], Rep (PDB 1uaa) [Korolev et al., 1997], UvrD (PDB 2is6) [Lee and Yang, 2006] and Rad54 (PDB ID 1z63), which superimpose ~ 200 - 250 C_{α} residues with ~ 3 - 3.5 Å rmsd on HsdR helicase domains 1 and 2, the HsdR helicase cleft accommodates double-stranded B-form DNA without clashes with the minor groove facing the protein

(Fig. 11-4; 11-5). By analogy with those cocrystal structures, DNA in the HsdR cleft is presumably transported toward the viewer in the orientation of **Figure 11-1**.

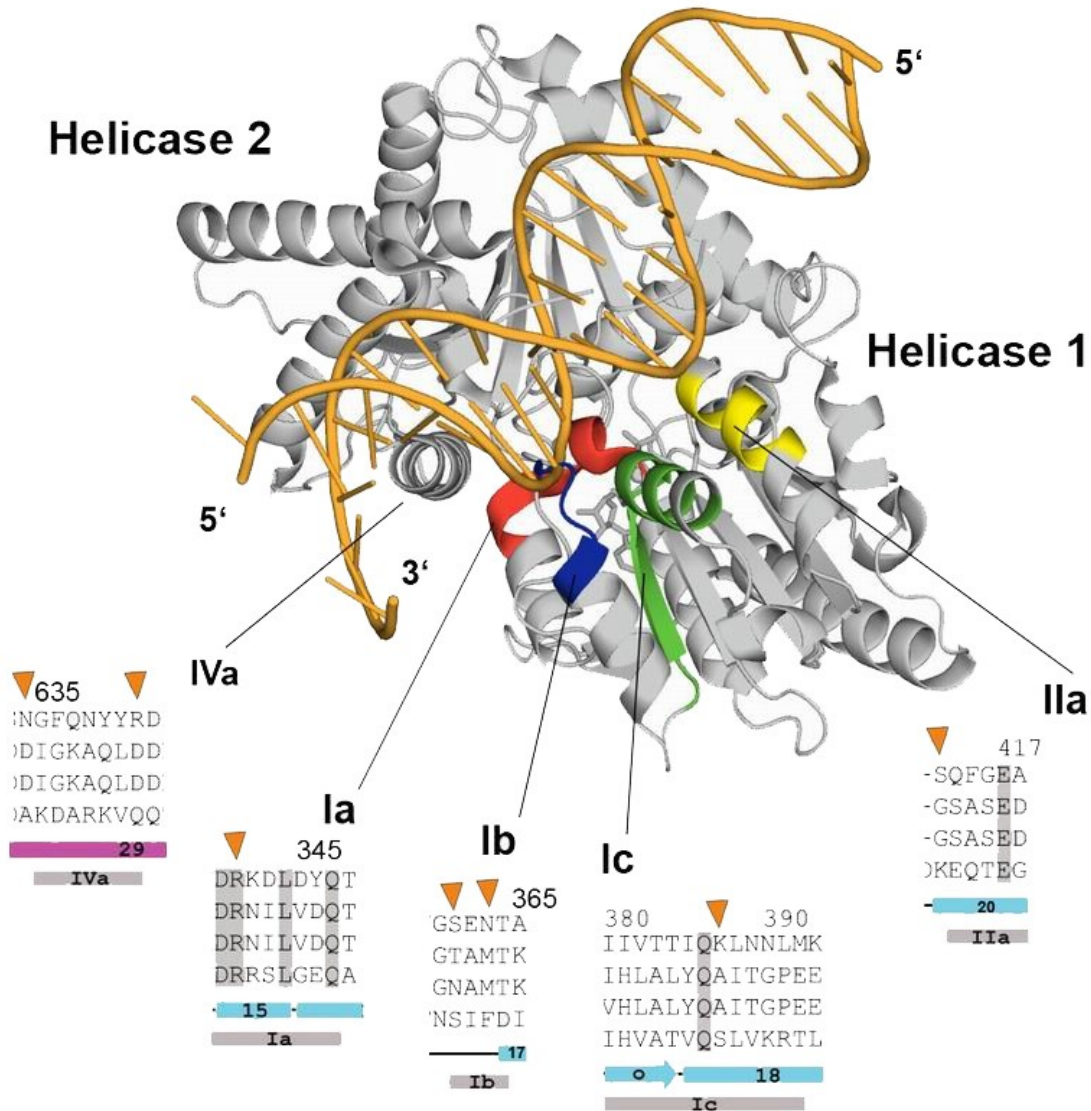


Fig. 11-5 HsdR-dsDNA complex model. The helicase domains are shown with the dsDNA been modeled in the cleft between them. Conserved motifs, distributed over the HsdR represented as a colored parts of a molecule. Corresponding alignment over the different families of type I R-M enzymes shown for each motif. Rows in alignment correspond to the alignment on **Fig. 11-2**. Residues potentially implemented in DNA binding are marked with triangles.

Sequence motifs identify the HsdR motor subunit as belonging to the so-called superfamily 2 (SF2) helicases, which include enzymes with true unwinding activity as

well as double-strand translocases that do not separate duplex strands. EcoR124I translocates in the 3' to 5' direction on one strand of the duplex without promoting extensive strand separation [Stanley et al., 2006], though local opening of the duplex in or near the enzyme complex has been suggested [van Noort et al., 2004]. The HsdR helicase cleft presents residues of motifs Ia, b, c, and II of domain 1 and motif IVa of domain 2, similarly to the residues and motifs involved in DNA contacts in related SF2 helicases cocrystallized with nucleic acids [reviewed by Dürr et al., 2005]. Each HsdR helicase domain is predicted to contact one DNA strand. Helicase domain 1 would contact the strand that has been defined in the helicase/DNA cocrystals as the 5' to 3' strand, making similar contacts. The 3' to 5' strand would contact helix 29 in domain 2 *via* motif IVa residues Asn632 and Arg639 (**Figure 11-5**). Although all known helicase structures have the equivalent of HsdR helix 29 that is proposed to be involved directly in nucleic acid transport [Singleton et al., 2007], the predicted or observed DNA contacts vary due to variations among the structures in the position of helicase domain 2 relative to domain 1.

DEAD-box motifs have been implicated in ATP hydrolysis. A variant DEAD box (DECHR, motif II) is located at HsdR residues 408-412 in the strand p-helix 20 turn (**Figure 11-5**). The closest approach of HsdR motif II to ATP (**Figure 11-6**) is made by Glu409 (4.85 Å from O1ε to ATP O1γ) This distance is similar for the corresponding Glu residues of other DEAD-box helicases that have a bridging water molecule between Glu O1ε and ATP O1γ [Sengoku et al., 2006]. Although in HsdR the available space is adequate to accommodate a water molecule, no density is detected, nor could a water occupy the same position as in the other structures because the position of the Glu409

carboxylate is altered by interaction with the ϵ -amino group of Lys313. Due to the absence of the water molecule a small change in the Glu409 C β -C γ dihedral angle would enable direct hydrogen bonding by bringing O1 ϵ within ~ 3.4 Å of ATP O1 γ (Fig. 11-6).

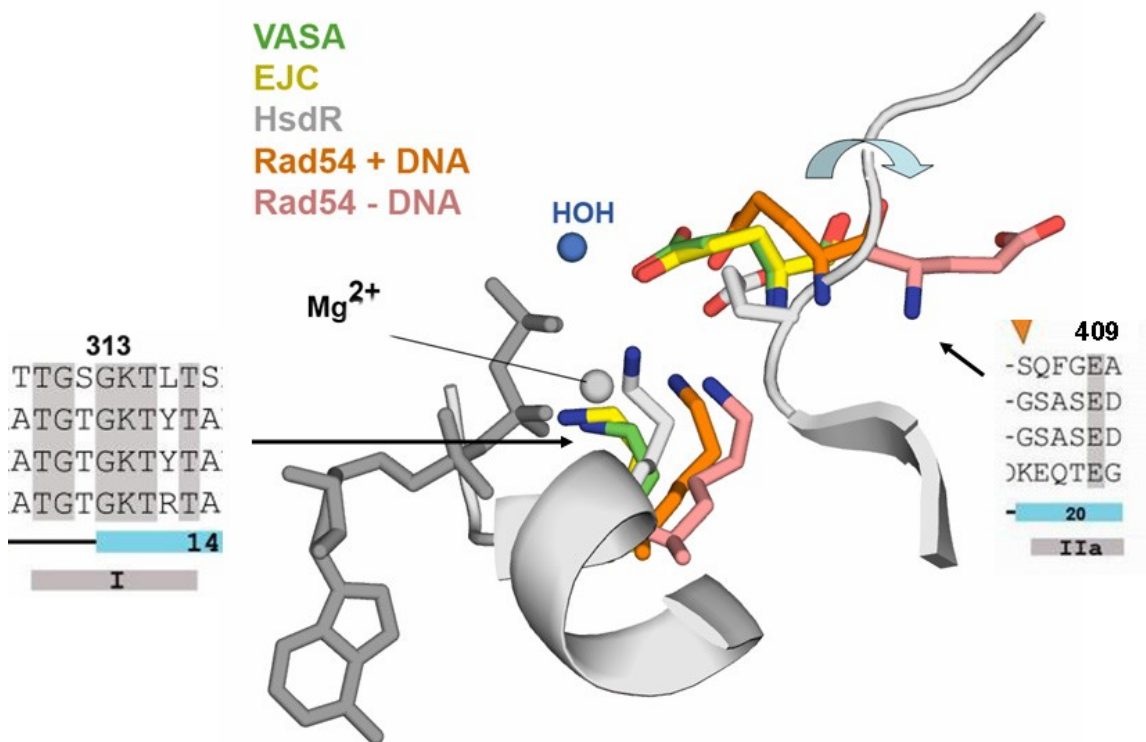


Fig. 11-6 The comparison of the ATP γ phosphate coordination differences among Vasa, EJC, HsdR and two binding forms of Rad54. The water molecule, proposed to be important in hydrolysis as shown from the analysis of Vasa and EJC proteins is absent in HsdR, which, in concord with the moving away Glu409 sidechain may be a part of an explanation of the ATP hydrolysis inhibition in the DNA-free ATP bound HsdR crystal form.

The adenine ring stacks on Arg273 and is surrounded by the other residues of motif Q. The N6 amino protons are close enough to the Val271 backbone carbonyl oxygen and the Gln276 sidechain amide to permit hydrogen bonding, as is N7 to Gln276. Residues of motifs I and VI are close enough to the oxygen atoms of all three ATP phosphates to form an extensive hydrogen-bond network. The ribose 3'OH proton and the Asp664 carboxylate of motif V likely form a charge-stabilized hydrogen bond. The 2'OH group is

distant from potential interactions; although this distance is similar in other helicases and has provided a structural explanation for their reported ability to use dATP to drive translocation *in vitro* [Lee and Yang, 2006], this ability is functionally unexplained, as dATP is not generally used as an energy source and is synthesized from ATP *in vivo*. The Mg²⁺ ion can be hexa-coordinated (distances ~2 Å) with octahedral symmetry by two water oxygens as axial ligands, and by phosphate oxygens O1β and O3γ, the hydroxyl oxygen of Thr314, and one water oxygen as equatorial ligands.

11.2.3 A Helical Domain for Subunit Assembly

The secondary and tertiary structure of the resolved proximal part of the HsdR helical domain through residue ~890 (helices 33-40) appeared by eye to resemble part of the helical domain of HsdM of EcoKI (**Figure 11-7a**). The unresolved distal part of the HsdR helical domain is predicted by the consensus of nine standard programs (not shown) to contain six additional helices spanning residues ~905 to 1005, suggesting that the potential similarity might extend over the entire domain. 3D superposition with HsdM confirms the similarity, although the statistical quality of the superposition (1.75 Å rmsd over 51 C_{alpha} atoms) is limited due to rigid-body shifts of the helices. The superposition predicts that the distal helices of the HsdR helical domain would fill the disordered gap without producing any clashes within the subunit. These results encouraged structural modeling of the HsdR helical domain. Energy minimization of the domain (residues 894 to 1038) was carried out to derive a complete tertiary structure encompassing the resolved and predicted helices. The crystal structure of EcoKI HsdM (PDB 2ar0) was not used directly as the template because it lacks helices 3 and 4 that are present in the resolved proximal part of the HsdR helical domain crystal structure as well as predicted

helix 9. Instead, the template chosen for 3D structural modeling was the energy-minimized, refined model structure of EcoR124I HsdM developed by Obarska et al., 2006 that includes those three helices. The resulting modeled structure of the complete HsdR helical domain (**Figure 11-7c**) indeed resembles HsdM, with the predicted helices enlarging the domain as suggested by the 3D superposition, making the domain more globular and more equal in size to the endonuclease domain. The modeled helical domain fills the space in the disordered gap (**Figure 11-7d**), making the motor subunit slightly larger, more symmetrical, and $\sim 10^\circ$ out of plane due to a slight shift in its center of mass. measurements on isolated HsdR. The maximum dimension determined from small-angle neutron scattering is ~ 100 Å; sedimentation velocity yields radius of gyration 34 Å and frictional coefficient 1.21 [Obarska-Kosinska et al., 2008]. The values calculated for modeled HsdR are maximum dimension ~ 90 Å, radius of gyration ~ 30 Å, and ratio of max/min dimensions ~ 1.3 . These small deviations between calculated and experimental measures would be consistent with a helical domain that is partially disordered and/or poorly accommodated in the crystal lattice. Indeed, the modeled motor subunit cannot be fit into the lattice due to steric clash with adjacent subunits (not shown). A nearby solvent channel contains very weak, discontinuous electron density that could not be interpreted even using the fully-modeled subunit as a search model; no other electron density in the map is unaccounted for, suggesting this uninterpretable density arises from the unresolved distal helical domain. The original report of HsdR purification found that only a minor species had the SDS gel migration expected for intact protein and the major species migrated faster [Janscak et al., 1996]. The presence of only two discrete species differing by ~ 7.4 kDa suggests that a labile region for endoproteolysis was located approximately 70 residues from the C-terminus around residue 970. The dimensions and

chemical character of the surface grooves of the fully-modeled motor subunit make a striking match to B-form DNA in a continuous path down the 'front' (as viewed in **Figure 11-1**) of the planar domain array. A positively charged surface groove proceeds from the

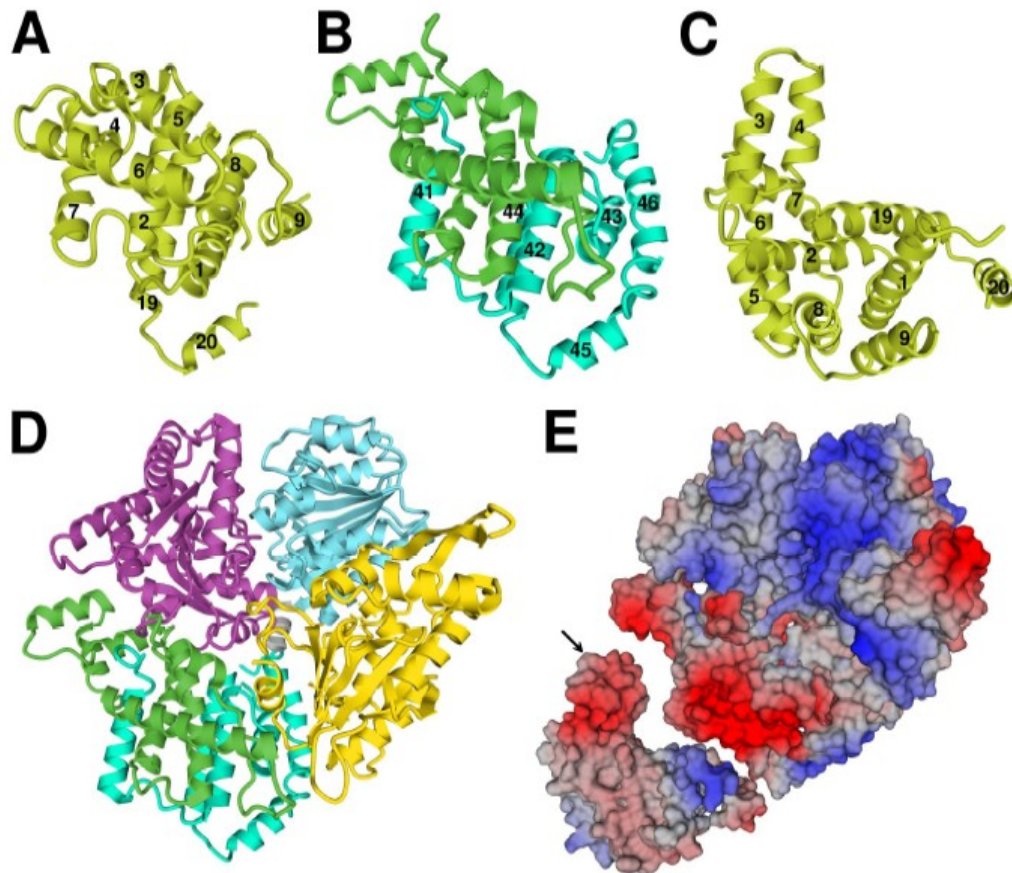


Fig. 11-7 Complementary helical domains of R and M subunits. (a) Model structure of EcoR124I HsdM helical domain. Model of Obarska et al., 2006 based on EcoKI HsdM crystal structure (PDB ID 2ar0), in viewpoint matching the HsdR helical domain on Figure 11-1. (b) Energy-minimized modeled HsdR helical domain. Viewpoint of panel a. Predicted helices are numbered. (c) EcoR124I HsdM model structure. Same as panel a but viewpoint is that of panel e. (d) Fully-modeled HsdR subunit with energy-minimized helical domain. Viewpoint of Figure 11-1. (e) HsdM/HsdR helical domain interface. Electrostatic potential surfaces of EcoR124I HsdM helical domain model (left), with orientation as in panel c, and fully-modeled HsdR subunit (right), with orientation as in panel d. The negative lobe at the top of the HsdM helical domain lies out of plan in this view, and would interact with a region of positive charge lying behind HsdR, as indicated by the arrow.

helicase cleft (**Figure 11-8a,b**), continues between the helical and endonuclease domains where it passes over the cleavage site recessed slightly from the surface (**Figure 11-8c,d**), motor subunit due to the extensive buried surface area of the domain assembly. DNA is

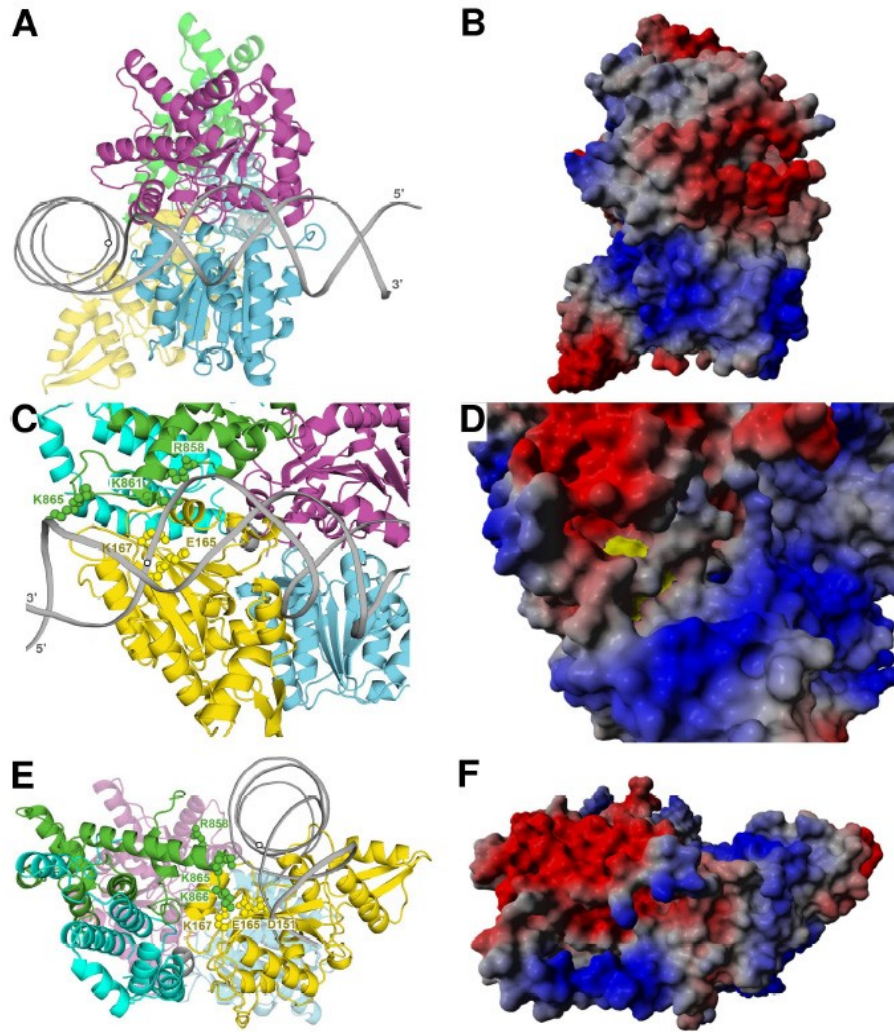


Fig. 11-8 Clefts and grooves accommodate DNA. Three rotated views are shown in ribbon models (left, with docked DNA) and corresponding surface electrostatic potential models (right) to display the continuous, complementary path proposed for DNA around the subunit. B-form DNA is represented by backbone ribbons, with 5' and 3' end labels defined by the directionality of strands contacting the motifs of helicase domains 1 and 2. An open circle on the 5' to 3' strand marks the DNA phosphate closest to the endonuclease active site. Positive surface electrostatic potential is represented in blue and negative potential in red, and active-site residues (Asp151, Lys165, Glu167) are represented in yellow space-filling spheres in all three panels though they are out of view in panels b and f. **(a,b) Helicase domain cleft.** Viewpoint is downward from the top of Figure 11-1, with front of subunit on the left. DNA is bent by $\sim 80^\circ$ as it leaves the

helicase cleft along the positively charged groove down the front of the subunit. **(c,d) Cleavage site groove.** The view is of the front of the subunit, with its top just beyond the right edge. Active-site (yellow) and exit clamp (green) residues lining the groove between endonuclease and helical domains are represented as space-filling spheres in domain colors and numbered in one-letter code. Some of the positively charged residues of the exit clamp are recessed below the negatively charged potential surface in this view. **(e,f). Motor subunit exit clamp.** Viewpoint is from left of panel c, with the front of the subunit facing upward. The recessed positive residues of the exit clamp are visible in this view.

likely to follow the shortest path along the groove because footprinting with the motor subunit or the enzyme complex shows continuous protection [Powell et al., 1998]. To make the shortest direct path along the surface groove, the DNA would have to bend (by approximately 80° deflection from linear) shortly after leaving the helicase cleft, as indicated in **Fig. 11-8**.

III. DISCUSSION

12. 180° ROTATION OF HELICASE DOMAIN 2

Among solved structures the distance and angle between the two helicase domains is quite variable, although the functional relevance of the differences has not been clear. The most extreme rotated state of the helicase domains has been observed among other SF2 translocases. Rad54 is the translocase most similar to HsdR in which helicase domains are crystallized in presence and absence of DNA, and its helicase 2 domain is rotated in both structures by $\sim 180^\circ$ from the position shown in **Fig. 11-1** [Dürr et al., 2005]. However, recent FRET analysis of the helicase domains of *S. solfataricus* Rad54 [Lewis et al., 2008] indicates that donor-acceptor distances large enough to represent 180° rotation of domain 2 are detected in solution only in the absence of DNA, suggesting the

180°-rotated state is not sampled during translocation. The completeness of the HsdR motor subunit structure model offers an opportunity to assess the structural compatibility and possible functional relevance of this rotation in the context of an intact, closely

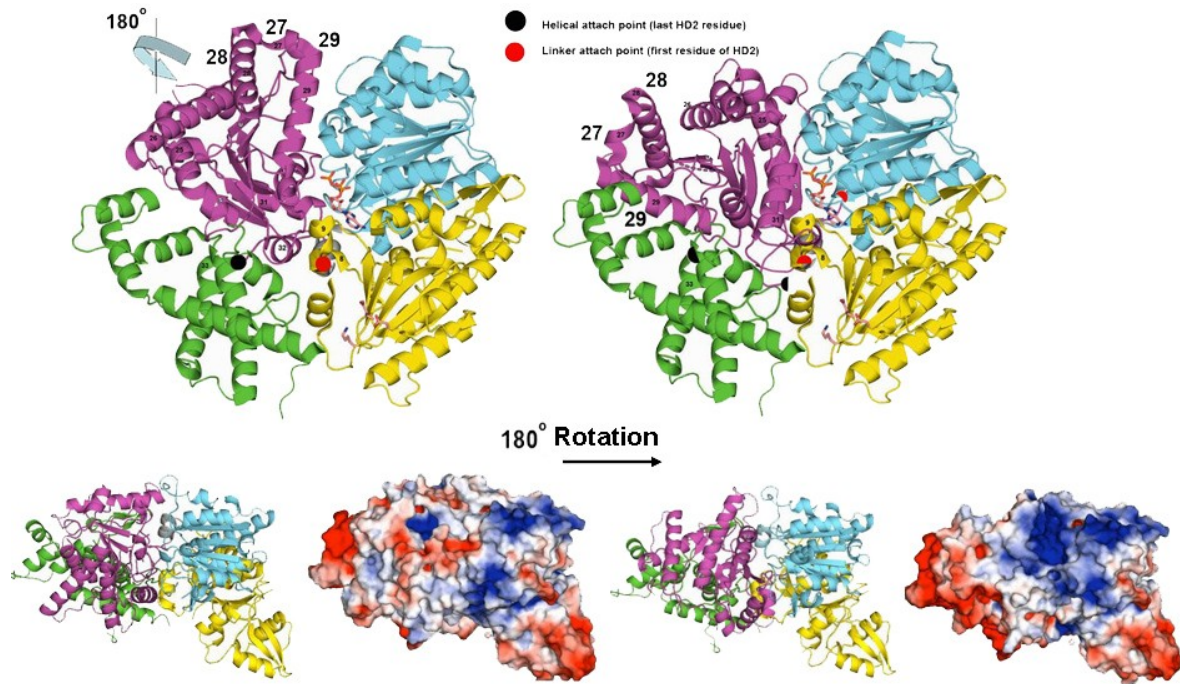


Fig. 12-1 The 180° degree rotation of the helicase domain 2 relative to the helicase domain 1, The rotation has been modeled based on the crystal structure model of the Rad54/dsDNA remodeling factor. The modeled big dislocation is not restricted by the accessories domains of HsdR suggesting a possibility of such transformation. Surface potentials, corresponding to two states: rotated and non-rotated shown in the lower half of the figure.

packed domain array rather than in isolated helicase domains. The HsdR helicase 2 domain structure model was excised, overlaid with the 180°-rotated state of Rad54, and replaced into the planar array (**Fig 12-1; 12-2**). In the intact subunit this rotation would require extensible chain segments of ~8 Å (equivalent to ~3 fully-extended residues) on the helicase 1 domain side and ~28 Å (~8 residues) on the helical domain side, which could be provided by unraveling short helices 22 and 32 flanking domain 2; interdomain

linker helix 22 is not fully helical in any other helicase structure and 9 into steric clash that would presumably affect the contact of ATP to Lys220 as well. In the HsdR subunit with fully-modeled helical domain no additional adjustments would be required to accommodate the 180°-rotated state (not shown). Thus, although steric clashes along its extensive interdomain interfaces might have precluded 180° rotation of HsdR helicase domain 2, the rotated state can apparently be accommodated in the motor subunit with only modest structural adaptations, but ATP contacts would be profoundly altered. Electrostatic potential surfaces are broader and more uniformly positive in the rotated

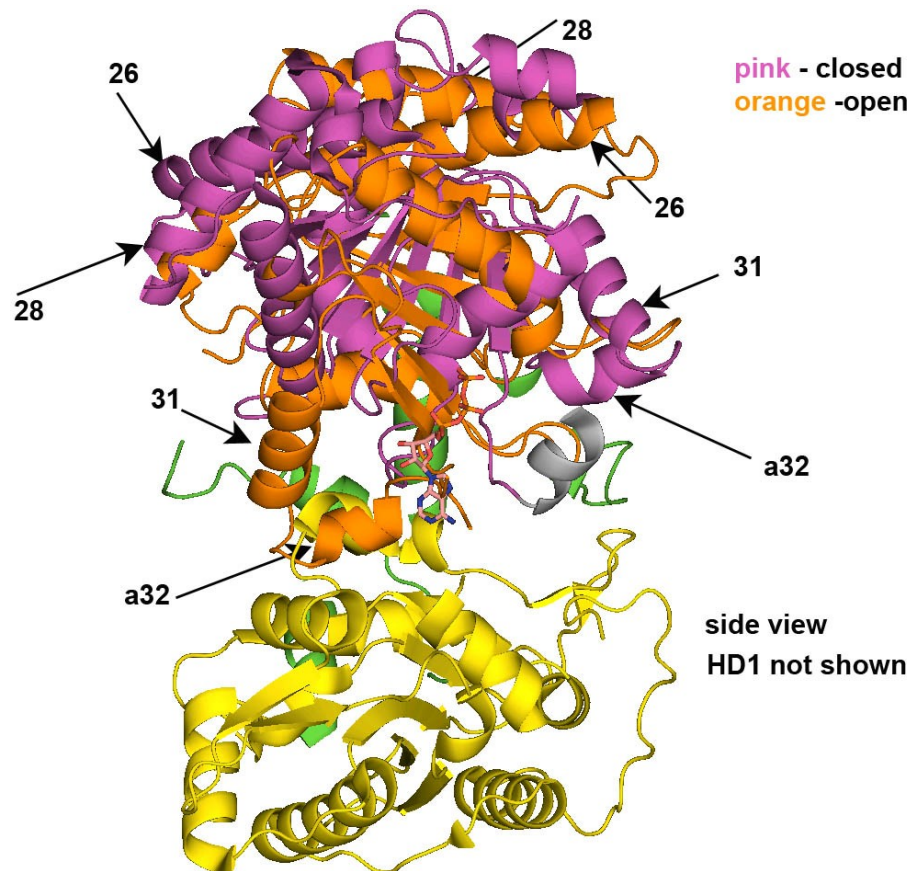


Fig. 12-2 The track of the helicase 2 domain rotation. Domain colored pink is the state like in HsdR crystal structure, representing a closed conformation while the one in orange is a Rad54 derived. The helices shown for the simplicity to follow the rotational dislocation. Helicase I domain is not shown. The Helical and Endonuclease domains shown only partly and color coded according to Fig. 11-1.

state of HsdR (**Figure 12-1**), presumably offering higher affinity than the unrotated state. These features recall the eukaryotic exon-junction complex, a persistent ribonucleoprotein assembly for nuclear export and translation, in which the ATPase activity of its DEAD-box helicase is proposed to be inhibited due to the tight grip of the complex on its mRNA target [Bono et al., 2006]. 180° rotation of HsdR helicase domain 2 could presumably accomplish both ATPase inhibition and increased target affinity at once. As well, the 180°-rotated HsdR model brings helices 27 and 28 lying just before helix 29 of helicase domain 2 close to the protruding knob formed by helices 36, 37, and 38 of the helical domain.

13. STAGES OF A ROTARY-INCHWORM STEPPING

FRET experiments with Rad54 in which the translocation catalytic cycle is staged using ATP, ADP, a non-hydrolysable ATP analog, or a transition-state mimic [Lewis et al., 2008] indicate that all stages sample a similar range of donor-acceptor distances. However, no distances compatible with 180° rotation are sampled during translocation despite the fact that crystals display 180° rotation in both presence and absence of DNA [Dürr et al., 2005]. Another example that suggests helicase domain orientation in crystals may be insensitive to ligand-occupancy states is UvrB, where structures with and without ATP bound superimpose even their sidechains with overall 0.7 Å rmsd [Theis et al., 1999]. Conversely, interdomain angles and distances can vary widely even among helicases with the same ligands bound. In aggregate such results offer no obvious structural correlation with stages of the translocation cycle, and they further suggest the possibility that in some ligand-occupancy states a range of conformations may be sampled. Because crystallization is not an equilibrium process it need not reflect the

distribution of conformations sampled in solution, but crystal structures presumably capture states that can be sampled by the domains. Thus, the available helicase crystal structures might be considered as a combinatorial group in which the range of interdomain angles/distances and cofactor states can be reassorted to represent known stages in translocation.

Excised HsdR domain 2 was superimposed with the corresponding domain of several helicase structures, all of which could be accommodated in the fully-modeled motor subunit without steric clash (not shown). Together with the HsdR structure itself, three other positions of domain 2 represented by crystal structures of RecG, NS3, and UvrB can be ordered into a sequence of motions, cofactor states, and DNA contacts that form a plausible series of translocation stages (**Figure 13-1c-j**). The first stage (**Figure 13-1c,d**) with no ATP bound is represented by the structure resulting from rotation of domain 2 by $\sim 45^\circ$ relative to crystalline HsdR about the same axis as the 180° rotation, as in RecG (PDB 1gm5). Upon DNA binding the 5'-3' strand would face helix 29 and the 3'-5' strand would face domain 1. The second stage (**Figure 13-1e,f**) with ATP bound is represented by crystalline HsdR with $\sim 0^\circ$ rotation. During rotation from the first to the second stage, DNA is transported by ~ 1 bp until the distal part of helix 29 slips off the duplex and contacts domain 1. The third stage (**Figure 13-1g,h**) with ADP bound is represented by the structure with $\sim -15^\circ$ rotation and a slight widening of the cleft by ~ 7 Å increase of the interdomain distance, as in NS3 (PDB 1a1v). This motion would accompany ATP hydrolysis and/or phosphate release, and could permit DNA release from domain 2. The fourth stage (**Figure 13-1i,j**) with no cofactor bound is represented by the structure with $\sim -5^\circ$ rotation and a further widening of the cleft by ~ 17 Å as in UvrB (PDB 1d9x).

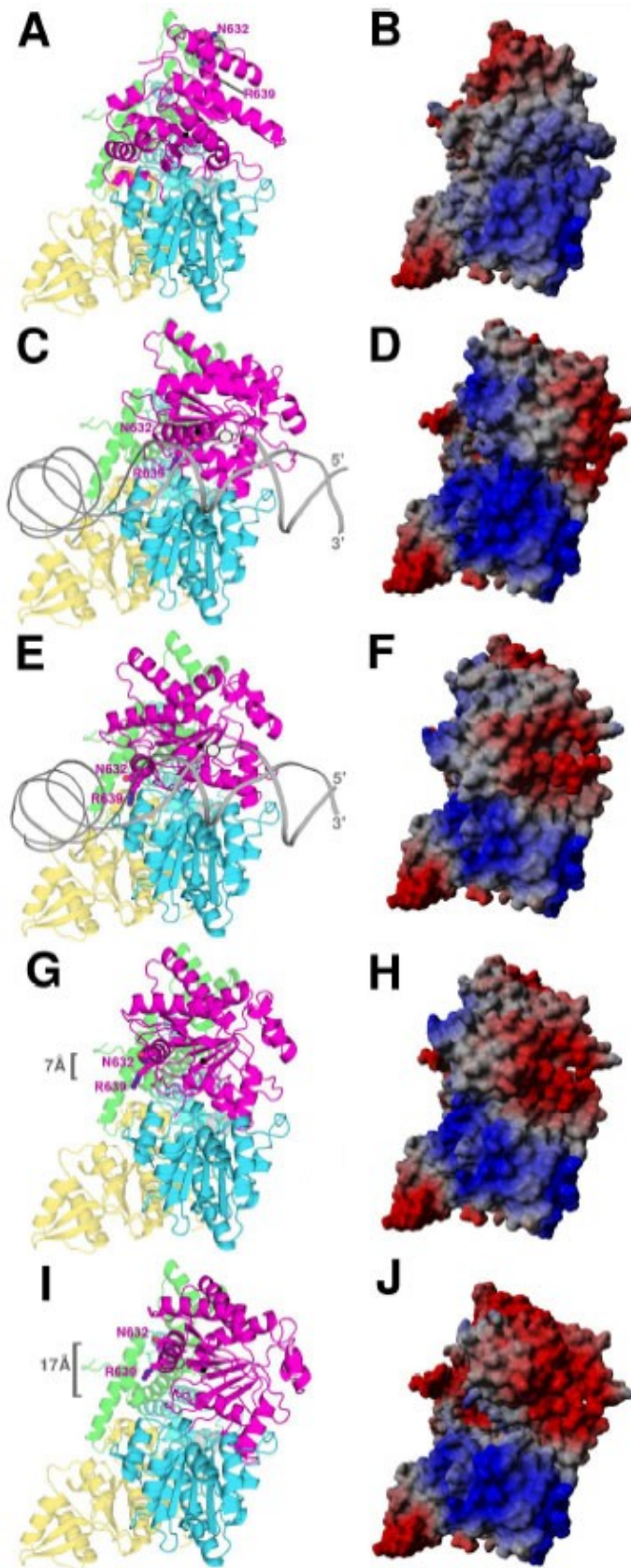


Fig. 13-1 Rotational states and translocation stages. The viewpoint is that of Figure 11-8. Left, ribbon models; right, electrostatic potential surfaces. In each pair of panels the helicase domain 2 structure was excised from the HsdR structure, superimposed on helicase domain 2 of the indicated crystal structure, and replaced in the HsdR structure in the new orientation as described in the text. The rotations of the domain can be followed by noting the changing positions of peripheral residues Asn632 and Arg639 (space-filling spheres in domain and atomic colors on helix 29) relative to the domain's center of mass (black dot); slight shifts in the center of mass are due to minor translational motions that accompany the rotations. Advancement of the DNA can be followed by noting the position of the open circle on the backbone ~ one turn from the marked 5' end. **(a,b).** $\sim 180^\circ$ rotation. This degree of rotation is modeled from the crystal structures of Rad54. In the translocation model this structure represents the stage before DNA binding. In this view helix 29 is nearly vertical at the top of the structure and only the nitrogen atoms (blue) of the Asn632 and Arg639 sidechains are visible. **(c,d)** $\sim 45^\circ$ rotation. Modeled from the crystal structure of RecG representing the stage with DNA bound before ATP binding. **(e,f)** $\sim 0^\circ$ rotation. This degree of domain rotation is observed in the HsdR structure itself representing the stage with ATP

bound and the DNA advanced by one bp. **(g,h) ~ -15° rotation and ~ 7Å separation.** Modeled from the structure of NS3 representing the stage with ADP bound after ATP hydrolysis. DNA is omitted for clarity. The two helicase domains have moved away from each other by ~7Å relative to panels a-f, as measured by the position of the Asn632 alpha carbon. **(i,j) ~ -5° rotation and ~ 17Å separation.** Modeled from the structure of UvrB with the domains far from each other (~17Å apart relative to panels a-f) representing the stage where ADP has been released, permitting resetting to the starting state of the translocation cycle shown in panel c.

The large interdomain distance after ADP release permits rotation back to the first stage represented by panel c. In agreement with the experimentally-determined step size [Seidel et al., 2008], the four stages result in translocation of the DNA by ~1 bp with consumption of one ATP, and resetting to the starting state.

The translocation stages traced out by the structures of **Fig. 13-1c-j** suggest an explicit molecular analogy with the mechanism used by a so-called rotary-inchworm motor. This device uses piezoelectric actuators to execute fast rotation of the motor shaft without sacrificing precision; an application that may be familiar is the stepping motor of a spectrophotometer monochromator. Three components cooperate to drive the motor shaft: a fixed one that provides an anchor, a cyclically attached one that drives the shaft, and a variable-distance connector between them. The molecular analog of the cyclically connected actuator would be helix 29; the distance regulator would be helicase domain 2 by virtue of its variable angles/distances; the anchor would be helicase domain 1 by virtue of its invariant contact with DNA and its covalent connection to domain 2; and the shaft is the DNA. Previously described macroscopic models for helicases and translocases have been referred to as inchworm models [Velankar et al., 1999; Hopfner and Michaelis, 2007]; the term rotary-inchworm motor has been applied as a mechanical analogy to the hexameric ring helicases [Massey et al., 2006]. To conclude, on **Fig. 13-2** the stages of the ATP-hydrolyzed coupled DNA translocation are shown:

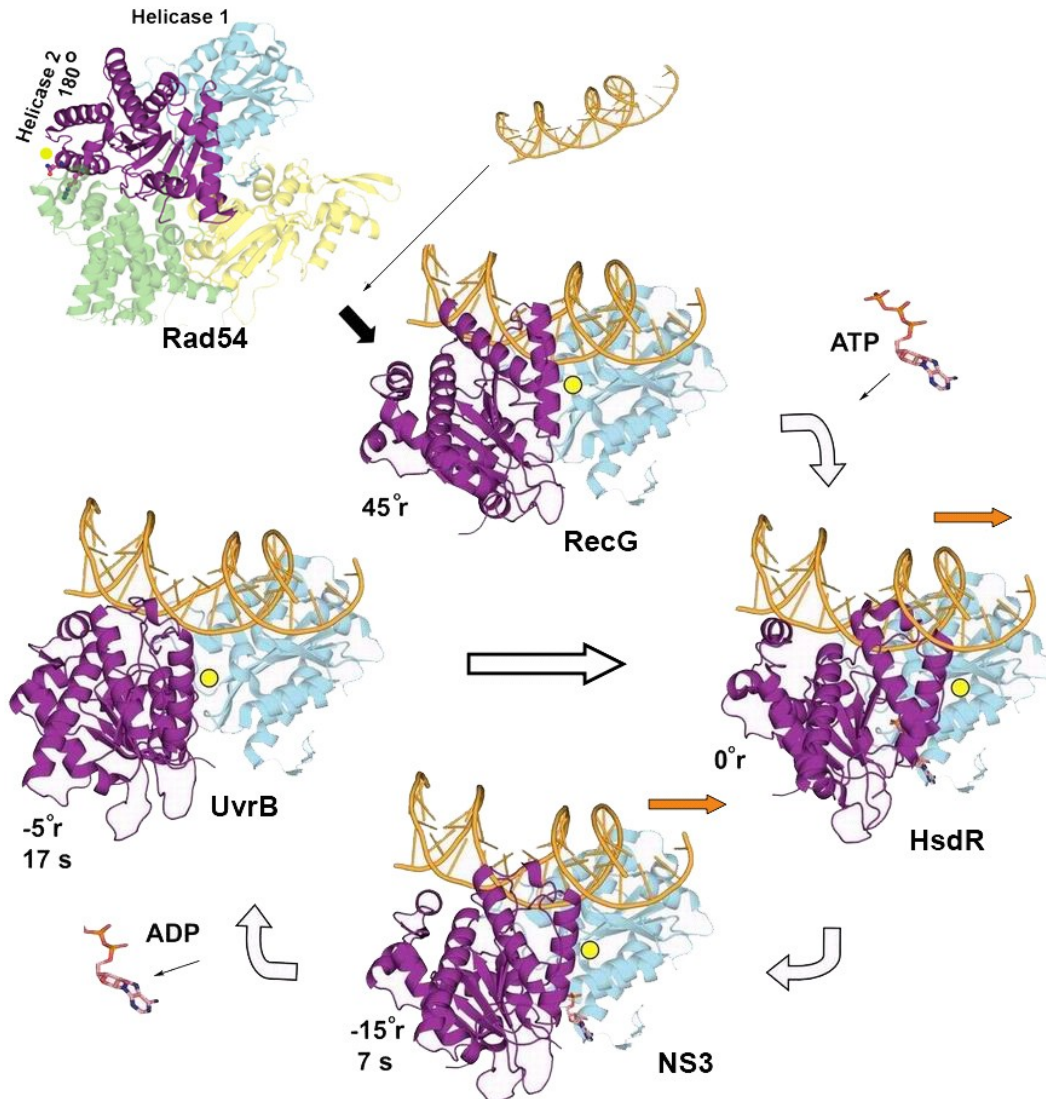


Fig 13-2. Stages of the ATP-dependant DNA translocation by HsdR subunit of EcoR124I enzyme. The helicae domain 2 is modeled according to its conformation occupancies in other SF2 proteins (shown by names at each stage). The first step is represented by the DNA-free state, as it was found in Rad54 studies. After dsDNA binding the helicase 2 domain undergoes a 45 degrees rotation and comes to the state observed in RecG helicase. ATP binding itself is sufficient to bring the helicase 2 domain in position like in HsdR structure which results in a direct DNA movement. As ATP is hydrolyzed and ADP is released (next two steps) the helicase 2 domain rearranges and returns to the initial DNA bound state after which the circle is repeated.

14. PENTAMERIC RESTRICTION COMPLEX

The vast body of mechanistic data available on the enzymatic activities of EcoR124I was used to constrain structural prediction of domain, subunit, and DNA interactions in the complete pentameric complex (**Fig. 14-1**). The starting point was a model structure of trimeric EcoR124I methyltransferase bound to duplex DNA [Obarska et al., 2006] developed from known structures of the related HsdM (PDB ID 2ar0) and HsdS (PDB ID 1ydx) subunits. Consistent with DNA footprinting [Taylor et al., 1993; Mernagh and Kneale, 1996] and hydrodynamic measurements [Callow et al., 2007] on the methyltransferase-DNA complex, that model shows HsdS approaching the DNA from one 'face' to bind the recognition sequence; two HsdM subunits approach from the opposite face to completely enclose the DNA over a length of ~60-80 Å (~20-24 bp). The catalytic domains of HsdM engage the recognition sequence and the bases to be methylated; the helical domains of HsdM are not directly involved in DNA binding.

The pentamer-DNA complex is seen as a dense globular object estimated (not shown) to cover ~80 bp in AFM images in the absence of cofactors [van Noort et al., 2004], indicating a compact organization of the enzyme with close proximity of HsdR to methyltransferase, consistent with the uninterrupted exonuclease III footprint [Powell et al., 1998]. The position of HsdR relative to methyltransferase is constrained by the known direction of DNA translocation toward the enzyme while engaged by the helicase domains [Bickle et al., 1978; Studier and Bandyopadhyay, 1988; Szczelkun et al., 1996; Ellis et al., 1999; Seidel et al., 2004]. This constraint suggests that one motor subunit is bound directly adjacent to, or slightly overlapping, each side of the methyltransferase, oriented with the helicase domains facing toward the flanking DNA, i.e., with the vertical

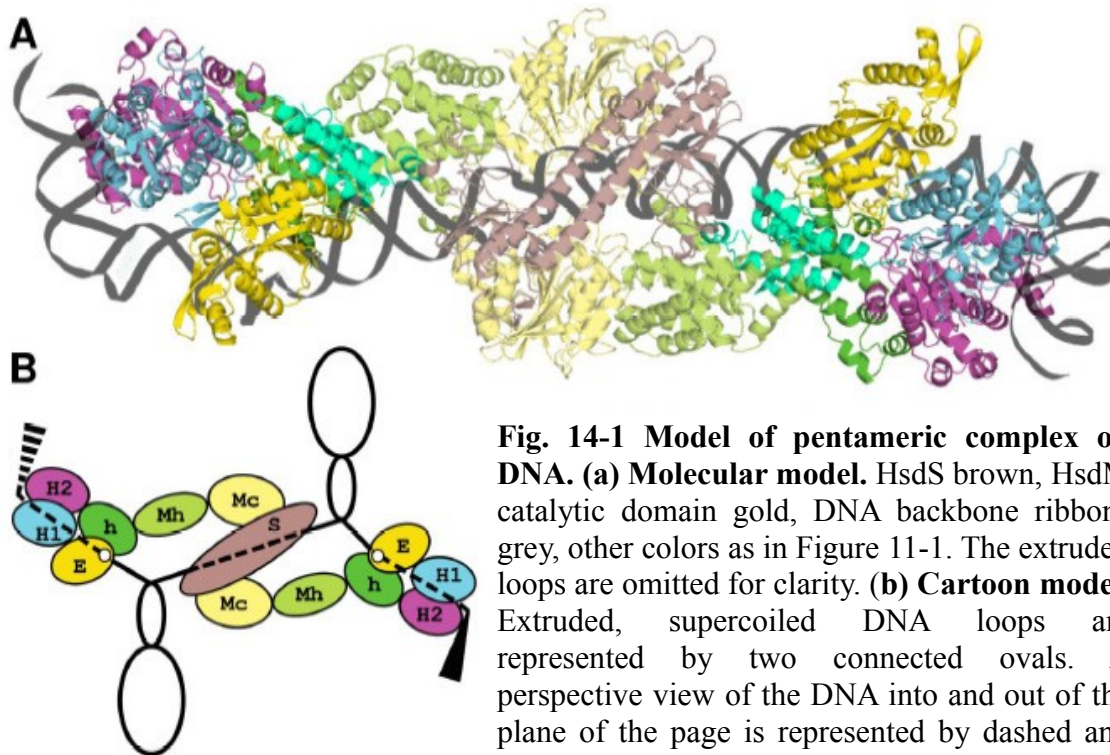


Fig. 14-1 Model of pentameric complex on DNA. (a) Molecular model. HsdS brown, HsdM catalytic domain gold, DNA backbone ribbons grey, other colors as in Figure 11-1. The extruded loops are omitted for clarity. **(b) Cartoon model.** Extruded, supercoiled DNA loops are represented by two connected ovals. A perspective view of the DNA into and out of the plane of the page is represented by dashed and solid triangles, respectively.

subunit axis of **Fig. 11-1** approximately parallel to the DNA long axis; this constraint does not specify an axial orientation. This placement would bring the helical domain of each motor subunit near the helical domain of each HsdM subunit, suggesting potential intersubunit contact between these two structurally similar domains. Consistent with this suggestion, removal of 155 residues from the C-terminus of EcoKI HsdR prevents assembly with methyltransferase [Davies et al., 1999], and substitution of A957 by Val is proposed to stabilize the HsdR/HsdM interface [Makovets et al., 2004]. The stability of the complex during purification and long after cleavage [Bickle et al., 1978; Yuan, 1981] suggests that subunit interactions are extensive. To achieve intimate contact the position of each HsdM helical domain was manually adjusted, which was enabled by two long, irregularly structured interdomain linkers that permit flexible orientation of the M subunit domains. Indeed, the mutual orientation of helical and catalytic domains in the EcoKI HsdM crystal structure is influenced by crystal packing [Obarska et al., 2006].

Furthermore, the dimensions of methyltransferase are drastically reduced upon DNA binding [Taylor et al, 1994], indicating that flexible domain attachment may be functionally significant. Taking account of all these factors, the EcoR124I HsdM helical domain was positioned to maximize the extent and symmetry of its contact with the helical domain of HsdR.

The resulting orientation indicates an excellent match of the shapes of the HsdR and HsdM helical domain surfaces (**Fig. 11-7e**). This placement of the helical domains is also consistent with patterns of surface charge covariance in the HsdM and HsdR subunits of other type I RMs. The EcoEI and EcoAI enzymes are the most closely related; their HsdM subunit sequences are > 90% identical, as are their HsdR sequences [Murray et al., 1993], and many of the differences are charge changes. Compared to EcoAI, the EcoEI enzyme has both a more negative HsdM and a more positive HsdR, with eleven more negative charges on HsdM and eleven more positive charges on HsdR. Most of these compensatory charge changes are on the helical domain, where they map to the surfaces of HsdM and HsdR that are proposed to interact in the pentameric complex model.

The axial orientation of HsdR domains becomes fixed by the choice of HsdM and HsdR helical domain interfaces. In the chosen orientation the surface grooves on the motor subunit align with the DNA where it emerges from within the methyltransferase complex. This placement fixes the end-to-end distance between helicase grooves of the two motor subunits at ~85 bp assuming B-form DNA, with one ~80° DNA bend at each end of the complex and with the second subunit ~90° out of helical phase with the first. The end-to-end DNA contour length calculated from the model is in exact agreement with AFM images of EcoR124I complexes [van Noort et al., 2004]. When one motor subunit joins the DNA-bound methyltransferase the AFM end-to-end length is shortened by ~8 nm,

and a second subunit shortens it by ~11 nm. When the complex model with two ~80° DNA bends that are out of plane by ~90° is viewed from 'above' as are AFM images, it would appear to be shortened by ~11 nm with two motor subunits bound, and by ~8 nm with one motor subunit bound.

The good agreement between dimensions of the complex model and AFM images for both one- and two-motor complexes suggests that the DNA is in B-form helical conformation on average through the complex. Thus, distortions around the bases to be methylated that are detected by hydroxyl radical footprinting of the methyltransferase complex [Mernagh and Kneale, 1996] are likely to be compensated locally. Short hairpin loops in or near the pentamer cannot be completely excluded, but the small bulge observed in AFM images [van Noort et al., 2004] might instead represent DNA that engages the helicase cleft upon ATP binding, concomitantly with acquisition of filter-binding competence at this step [Yuan 1981].

The model is also consistent with footprinting results for pentameric EcoKI- and EcoR124I-DNA complexes. Addition of two motor subunits lengthens the EcoKI methyltransferase footprint from ~28 bp to ~62 bp [Powell et al., 1998], a difference of ~125 Å, very similar to the summed dimensions of two HsdR subunits (130 Å) measured along the horizontal part of the DNA long axis in **Figure 14-1a**. Footprinting analysis of the EcoR124I methyltransferase is in agreement with the EcoKI results [Mernagh and Kneale, 1996], but the 56-bp duplex used is too short to engage the motor subunits and the EcoR124I footprint does not lengthen [Mernagh et al., 1998], suggesting that recruitment of motor subunits by protein-protein contact between HsdM and HsdR helical domains depends on DNA contact by HsdR. A cartoon model derived from

extensive biochemical data on the subunit and DNA interactions of EcoKI [Davies et al., 1999] is remarkably similar to **Figure 14-1b**.

15. DNA TRANSLOCATION AND CLEAVAGE

As helicase action pulls double-stranded DNA toward the stationary enzyme by tracking along the DNA helical pitch, negatively supercoiled DNA loops are extruded from within the bound pentamer [Seidel et al., 2004]; on circular or tethered linear DNAs positively supercoiled loops are also formed in the DNA ahead of the bound enzyme. The pentamer model suggests a probable site for DNA extrusion along a positively charged groove formed between symmetry-related positions on the interacting helical domains of HsdR and HsdM. On the HsdR side this groove is lined by Arg858 and 859, and Lys861, 865, 866, and 868, all at the C-terminal end of helix 39. The HsdM side of the groove presents Lys98 after helix 5, and Lys148 and 150 and Arg153 flanking helix 8. This groove acts as an exit clamp for the DNA as it passes into the extruded loop. The predicted close fit and strong attraction of the exit clamp for DNA are expected to favor maintenance of the duplex state and to confine accumulating negative supercoils to the growing loop. A tight grip on the DNA in its location just downstream of the endonuclease active site suggest the exit clamp could influence cleavage.

Possible positions for DNA at the endonuclease active site were modeled by comparison with cocrystal structures of dimeric type II restriction enzymes in complex with their cognate DNAs. Superposition of the common endonuclease cores of HsdR and the type II enzymes predicts that the 5' to 3' strand would be positioned over the HsdR active site with the major groove facing the protein. Unlike the type II case where a second subunit presents a symmetry-related active site, the 3' to 5' strand would have no direct access to

HsdR catalytic residues unless the 5' to 3' strand moves out of the way. This finding may be relevant to persistent but inconclusive reports for several type I RMs that more than one enzyme complex is required to complete one double-strand cleavage [Yuan, 1981]. Furthermore, EcoKI enzyme-DNA complexes can form dimers independently of translocation, ATP binding, or specific DNA recognition sites [Ellis et al., 1999; Berge et al., 2000], suggesting cooperation between pentamers. Attempts to model a dimeric complex of DNA-bound pentamers produced a range of possible arrangements. Based on absence of protein or DNA steric clash the most probable of these aligns the pentamers back-to-back (with 'front' being the view shown in **Figure 14-1a**) and with the two S subunits forming an X with each other in the view of **Figure 14-1a** (not shown). In this orientation the exposed surfaces of the helical domains of HsdR and HsdM subunits offer chemical complementarity of interacting surfaces. This arrangement would also bring into proximity the exit clamps and extruded DNA segments of the two pentamers. A putative role for the HsdM and HsdR helical domains in recruitment of one pentamer by another mirrors their role in recruitment of motor subunits by individual pentamers, including a dependence on bound DNA.

The shortest length of DNA between two HsdR endonuclease active sites of adjacent pentamers in such a dimer would be ~50 bp. Cleavage of circular DNAs containing a single recognition site produces a class of products (about 15% of the total) with 3' overhangs of ~45-50 nucleotides [Jindrova et al., 2005] that could be accounted for by this model. However, this length also corresponds to the distance between endonuclease active sites in one pentamer, consistent with cleavage of one strand by each HsdR subunit following translocation of the entire circular substrate by a single pentamer. Available

biochemical data do not clearly favor one interpretation of this class of products, nor do they define the enzyme-DNA stoichiometry required for cleavage.

The remaining ~85% of cleavage products display a bewildering variety of DNA termini, with 5' or 3' overhangs ranging in length from a couple of nucleotides to more than 50 and no clearly preferred lengths; only blunt ends can be ruled out as cleavage products. The largest class of cleavage products, approximately 45% of the total, has 5' overhangs of 2 to 5 nucleotides. Such short overhangs cannot easily be explained if one motor subunit cleaves only one strand, and instead suggests that each subunit may make a proximal cut in the second strand. The heterogeneity of these overhangs could result from precise cleavage of the first strand, presumably the 5' to 3' strand, followed by its displacement from the active site, with the increased flexibility of the nicked DNA permitting less precise cleavage of the second strand. Compared with the heterogeneity of these shorter products, the fact that as many as ~15% of products have ~50-nucleotide 3' overhangs suggests a substantial frequency of events in which cleavage of the second strand proximal to the nick is slow, consistent with poor access to the same active site. The remaining ~40% of products is distributed nearly evenly among overhangs of ~1 to 40 nucleotides, ~90% of them on the 3' side. The mechanistic significance of this large fraction of heterogeneous 3' overhangs is not immediately clear, but it could be related to the protracted ATPase activity following cleavage [Eskin and Linn, 1972; Yuan, 1981] if two motor subunits engage in a tug-of-war on a nicked substrate.

Although EcoR124I cuts circular DNAs bearing only one recognition site, a minimum of two sites is required on linear DNAs for efficient cleavage, which occurs midway between them [Studier and Bandyopadhyay, 1988]. Midway cleavage between pentamers is consistent with reports that several distinct kinds of impediments to translocation can

trigger cleavage [Szczelkun et al., 1996; Janscak et al., 1999; Firman and Szczelkun, 2000]. DNA supercoiling also affects translocation rate, processivity, and cleavage [Janscak et al., 1996; Seidel et al., 2004]. Together these results imply a cleavage-triggering mechanism that depends on DNA tension, perhaps simply by completion of translocation, the common endpoint of helicase activity on both circular DNAs with single sites and linear DNAs with multiple sites or impediments. Although midway cleavage has been assumed to occur after completion of translocation by two independent pentamers pulling the DNA between them in opposite directions, it appears to be equally compatible with a model in which dimerization precedes the completion of translocation. Relatively strong nonspecific binding may account for reports that excess enzyme promotes cleavage of linear DNAs with single recognition sites [Murray et al., 1973; Yuan, 1981]; nonspecific binding could also permit a second pentamer to promote midway cleavage on circular DNAs with single recognition sites. Careful biochemical analysis will be required to clarify the role of cooperation between pentamers in translocation, cleavage triggering, and catalysis.

CONCLUDING REMARKS AND OTHER PROJECTS

1. The native and the selenomethionine-labeled HsdR subunit of EcoR124I R-M complex was expressed and purified to homogeneity for the use in structural studies.
2. The purified proteins were crystallized (Lapkouski et al., *Acta. Cryst. F*, 2007), leading to the successful X-ray data collection.
3. The first ever X-ray crystal structure of HsdR protein was determined at 2.6 Å by selenomethionine SAD experiment from the derivative crystal of HsdR (PDB ID 2w00).
4. The residues involved in ATP coordination have been located. The model of the HsdR-dsDNA complex has been build based on the solved structure and available structural and biochemical data. The residues involved in DNA binding have been located based on the model.
5. The model of the pentameric EcoR124I-DNA complex was generated using mechanistic data of the solved HsdR protein and structures of related methylase and specificity subunits. The results suggest how rearrangements and cooperation of subunits, domains, and substrates initiate and stabilize the translocating complex as it tracks on DNA (manuscript has been recently accepted to *Nat. Str. Mol. Biol.*).
6. The reliable model of the molecular coupling of the ATP-dependent dsDNA motion has been proposed which could be applied to similar enzymes from type I restriction family and beyond.

Apart of my main project I took part in other projects shortly described below:

1) One of the project I was involved in is focused on the structure based studies of the Haloalkane dehalogenases, which are members of the α/β hydrolase fold family and catalyze the hydrolytic conversion of a broad spectrum of hydrocarbons to the corresponding alcohols. Haloalkane dehalogenases are considered to be important biocatalysts in bioremediation applications to decontaminate contaminated environments. We crystallized and collected native data sets for dehalogenases DhaA12 (PDB ID 2v9z), DhaA04, DhaA14 and DhaA15 to resolutions of 3.0, 1.30, 0.95 and 1.15 Å, respectively (paper is attached).

2) The second project I participated in is dealing with the structural studies on the extrinsic proteins at the luminal surface of the thylakoid membrane which are part of the higher plants Photosystem II: PsbO, PsbP and PsbQ. We collected 1.98 Å data for the native PsbP protein at ESRF (European Synchrotron Radiation Facility) in Grenoble, and deposited the coordinates under the PDB code 2vu4 (crystallization manuscript is attached; structure to be published).

3) We have recently finalized the work on the flavoprotein WrbA from *Escherichia coli*, which is considered to be the prototype of a new family of multimeric flavodoxin-like proteins that are implicated in cell protection against oxidative stress. Structure of the unliganded form of WrbA recombinant protein was determined at 1.8 Å resolution and deposited under PDB ID 2rg1 and to be published.

REFERENCES

- Abadjieva, A., Webb, M., Patel, J., Zinkevich, V., and Firman, K. Deletions within the DNA recognition subunit of M.EcoR124I that identify a region involved in protein-protein interactions between HsdS and HsdM. *J. Mol. Biol.* 241(1), 35-43 (1994).
- Aravind, L., Makarova, K.S., and Koonin, E.V. Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucl. Acids Res.* 28, 3417–3432 (2000).
- Arber, W., and Dussoix, D. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.* 5, 18–36 (1962).
- Asherie, N. Protein crystallization and phase diagrams. *Methods.* 34(3), 266-272 (2004).
- Bannister, D., Glover, S.W. Restriction and modification of bacteriophages by R⁺ strains of *Escherichia coli* K12. *Biochem. Biophys. Res. Commun.* 27, 30(6), 735-738 (1968).
- Barcus, V.A., Titheradge, A.J.B., and Murray, N.E. The diversity of alleles at the hsd locus in natural populations of *Escherichia coli*. *Genetics.* 140, 1187-1197 (1995).
- Bernstein, D.A., Zittel, M.C., and Keck, J.L. High-resolution structure of the E.coli RecQ helicase catalytic core. *EMBO J.* 22, 4910-4921 (2003).
- Berge, T., Ellis, D.J., Dryden, D.T.F., Edwardson, J.M., and Henderson, R.M. Translocation-independent dimerization of the EcoKI endonuclease visualized by atomic force microscopy. *Biophys. J.* 79, 479-484 (2000).
- Bertani, G., and Weigle, J.J. Host-controlled variation in bacterial viruses. *J. Bacteriol.* 65, 113–121 (1953).
- Bianco, P.R., Hurley, E.M. The type I restriction endonuclease EcoR124I, couples ATP hydrolysis to bidirectional DNA translocation. *J. Mol. Biol.* 352(4), 837-859 (2005).
- Bickle, T.A. The ATP-dependent restriction enzymes. In *Nucleases* (Linn, S. M., Lloyd, R. S. and Roberts, R. J., eds), 2nd edit., pp. 35±88, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1993).
- Bickle, T.A., Brack, C., and Yuan, R. ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli* K-12. *Proc. Natl Acad. Sci. USA*, 75, 3099–3103 (1978).
- Bono, F., Ebert, J., Lorentzen, E., and Conti, E. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126, 713-725 (2006).
- Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. Crystallography and NMR

system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* D54, 905–921 (1998).

Bujnicki, J.M., and Rychlewski, L. Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles, *J. Mol. Microbiol. Biotechnol.* 3, 69–72 (2001).

Calisto, R.M., Pich, O.Q., Pinol, J., Fita, I., Querol, E., Carpena, X. Crystal structure of a putative type I restriction-modification S subunit from *Mycoplasma genitalium*. *J. Mol. Biol.* 351, 749-762 (2005).

Callow, P., Sukhodub, A., Taylor, J.E., and Kneale, G.G. Shape and subunit organisation of the DNA methyltransferase M.AhdI by small-angle neutron scattering, *J. Mol. Biol.* 369, 177–185 (2007).

Caruthers, J.M., and McKay, D.B. Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* 12, 123–133 (2002).

Chin, V., Valinluck, V., Magaki, S., and Ryu, J. KpnBI is the prototype of a new family (IE) of bacterial type I restriction-modification system. *Nucleic Acids Res.* 8, 32(18), 138 (2004).

Cooper, L.P. and Dryden, D.T.F. The domains of a type I DNA methyltransferase. Interactions and role in recognition of DNA methylation. *J. Mol. Biol.* 236, 1011–1021 (1994).

Cowan, G.M., Gann, A.A., and Murray, N.E. Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell* 13, 56(1), 103-109 (1989).

Cowtan, K. Jnt CCP4/ESF-EACBM Newsl. *Protein Crystallogr.* 31, 34-38 (1994).

Cuff, J.A., and Barton, G.J. Application of enhanced multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* 40, 502-511 (1999).

Davies, G.P., Kemp, P., Molineux, I.J., and Murray, N.E. The DNA translocation and ATPase activities of restriction-deficient mutants of EcoKI. *J. Mol. Biol.* 292, 787–796 (1999b).

Davies, G.P., Martin, I., Sturrock, S.S., Cronshaw, A., Murray, N.E. and Dryden, D.T. On the structure and operation of Type I DNA restriction enzymes. *J. Mol. Biol.* 290, 565–579 (1999).

Dreier, J., MacWilliams, M.P. and Bickle, T.A. DNA cleavage by the type IC restriction-modification enzyme EcoR124II. *J. Mol. Biol.* 264, 722-733 (1996).

Dryden, D.T.F., Cooper, L.P. and Murray, N.E. Purification and characterization of the methyltransferase from the type I restriction and modification system of *Escherichia coli* K12. *J. Biol. Chem.* 268, 13228-13236 (1993).

Dryden, D.T.F., Cooper, L.P., Thorpe, P.H. and Byron, O. The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications. *Biochemistry.* 36, 1065-1076 (1997).

Dürr, H., Körner, C., Müller, M., Hickmann, V., and Hopfner, K.P. X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* 121, 363–373 (2005).

Ellis, D.J., Dryden, D.T., Berge, T., Edwardson, J.M., and Henderson, R.M. Direct observation of DNA translocation and cleavage by the EcoKI endonuclease using atomic force microscopy. *Nat. Struct. Biol.* 6, 15-17 (1999).

Emsley, P., and Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* D60, 2126-2132 (2004).

Endlich, B., and Linn, S. The DNA restriction endonuclease of *Escherichia coli* B. II. Further studies of the structure of DNA intermediates and products *J. Biol. Chem.* 260, 5729–5738 (1985).

Eskin, B., and Linn, S. The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. II. Purification, subunit structure, and catalytic properties of the restriction endonuclease. *J. Biol. Chem.* 247, 6183-6191 (1972).

Firman, K., Creasey, W.A., Watson, G., Price, C., and Glover, S.W. Genetic and physical studies of restriction-deficient mutants of the Inc FIV plasmids R124 and R124/3. *Mol. Gen. Genet.* 191(1), 145-153 (1983).

Firman, K., and Szczelkun, M.D. Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. *EMBO J.* 19, 2094–2102 (2000).

Firman, K., Creasey, W.A., Watson, G., Price, C. and Glover, S.W. Genetic and physical studies of restriction-deficient mutants of the Inc FIV plasmids R124 and R124/3. *Mol. Gen. Genet.* 191, 145-153 (1983).

Gill, S.C., and von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319-326 (1989).

Glover, S.W., Firman, K., Watson, G., Price, C., and Donaldson, S. The alternate expression of two restriction and modification systems. *Mol. Gen. Genet.* 190(1), 65-69 (1983).

Gorbalenya, A.E., and Koonin, E.V. Endonuclease (R) subunits of type-I and type-III restriction–modification enzymes contain a helicase-like domain. *FEBS Lett.* 291, 277–281 (1991).

Grune, T. Protein Crystallography. Part II, 2005. <http://shelx.uni-ac.gwdg.de>

Guex, N., and Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis.* 18, 2714-2723 (1997).

Hall, M.C., and Matson, S.W. Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* 34, 867–877 (1999).

Hedges, R.W., and Datta, N. R124, an ϕ R factor of a new compatibility class. *J Gen Microbiol.* 71(2), 403-405 (1972).

Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G., and Gibson, T.J. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* 22, 4673–4680 (1994).

Holubova, I., Vejsadová, S., Firman, K., and Weiserová, M. Cellular localization of Type I restriction-modification enzymes is family dependent. *Biochem. Biophys. Res. Commun.* 319(2), 375-380 (2004).

Holubova, I., Vejsadova, S., Weiserova, M., and Firman, K. Localization of the Type I restriction–modification enzyme EcoKI in the bacterial cell, *Biochem. Biophys. Res. Commun.* 270, 46–51 (2000).

Hopfner, K.P., and Michaelis, J. Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics. *Curr. Opin. Struct. Biol.* 17, 87–95 (2007).

Janscak, P., Abadjieva, A., and Firman, K. The type I restriction endonuclease R.EcoR124I: over-production and biochemical properties. *J. Mol. Biol.* 257(5), 977-991 (1996).

Janscak, P., and Bickle, T.A. DNA supercoiling during ATP-dependent DNA translocation by the type I restriction enzyme EcoAI. *J. Mol. Biol.* 295, 1089-1099 (2000).

Janscak, P., MacWilliams, M.P., Sandmeier, U., Nagaraja, V., and Bickle, T.A. DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J.* 18(9), 2638-2647 (1999a).

Janscak, P., Sandmeier, U., and Bickle, T.A. Single amino acid substitutions in the HsdR subunit of the type IB restriction enzyme EcoAI uncouple the DNA translocation and DNA cleavage activities of the enzyme. *Nucleic Acids Res.* 27(13), 2638-2643 (1999b).

Janscak, P., Dryden, D. T. F. and Firman, K. Analysis of the subunit assembly of the type IIC restriction-modification enzyme EcoR124I. *Nucl. Acids Res.* 26, 4439–4445 (1998).

Jindrova, E., Schmid-Nuoffer, S., Hamburger, F., Janscak, P. and Bickle, T.A. On the DNA cleavage mechanism of type I restriction enzymes. *Nucl. Acids Res.* 33, 1760-1766 (2005).

Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195-202 (1999).

Jung, J., and Lee, B. Protein structure alignment using environmental profiles. *Prot. Eng.* 13, 535-543 (2000).

Kim, J.-S., DeGiovanni, A., Jancarik, J., Adams, P.D., Yokota, H., Kim, R., and Kim, S.-H. Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc. Natl. Acad. Sci. U. S. A.* 102, 3248-3253 (2005).

Kim, J.L., Morgenstern, K.A., Griffith, J.P., Dwyer, M.D., Thomson, J.A., Murcko, M.A., Lin, C., and Caron, P.R., Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* 6, 89–100 (1998).

Kimball, M., and Linn, S. The release of oligonucleotides by the *Escherichia coli* B restriction endonuclease *Biochem. Biophys. Res. Commun.* 68, 585–591 (1976).

Kleywegt, G.J. Crystallographic refinement of ligand complexes. *Acta Crystallogr.* D63, 94-100 (2007).

Kneller, D.G., Cohen, F.E., and Langridge, R. Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.* 214, 171-182 (1990).

Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M., and Waksman, G. Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* 90, 635-647 (1997).

Kosinski, J., Feder, M., and Bujnicki, J.M. The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinformatics.* 6, 172 (2005),

Krieger, E., Darden, T., Nabuurs, S.B., Finkelstein, A., and Vriend, G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* 57, 678-683 (2004).

Kusano, K., Sakagami, K., Yokochi, T., Naito, T., Tokinaga, Y., Euda, E. and Kobayashi, I. A new type of illegitimate recombination is dependent on restriction and homologous interaction. *J. Bacteriol.* 179, 5380–5390 (1997).

La Fortelle, E., and Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction. *Meth. Enzymol.* 276, 472-494 (1997).

Lapkouski, M., Panjikar, S., Kuta Smatanova, I., and Csefalvay, E. Purification, crystallization and preliminary X-ray analysis of the HsdR subunit of the EcoR124I endonuclease from *Escherichia coli*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* F63, 582-585 (2007).

Lee, J., and Yang, W. UvrD Helicase Unwinds DNA One Base Pair at a Time by a Two-Part Power Stroke. *Cell* 127, 1349-1360 (2006).

Lewis, R., Dürr, H., Hopfner, K.P., and Michaelis, J. Conformational changes of a SWI2/SNF2 ATPase during its mechanochemical cycle. *Nucl. Acids Res.* 36, 1881-1890 (2008).

Linn, S., and Arber, W. Host specificity of DNA produced by *Escherichia coli*. X. In vitro restriction of phage fd replicative form. *Proc. Natl. Acad. Sci. USA.* 59, 1300–1306 (1968).

Makovets, S., Powell, L.M., Titheradge, A.J., Blakely, G.W., and Murray, N.E. Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease? *Mol. Microbiol.* 51, 135–147 (2004).

Maniatis, T., Fritsch, E. F. and Sambrook, J. *Molecular Cloning: A Laboratory Manual*. New York (1982).

Massey, T.H., Mercogliano, C.P., Yates, J., Sherratt, D.J., and Lowe, J. Double-stranded DNA translocation: Structure and mechanism of hexameric FstK. *Mol. Cell* 23, 457-469 (2006).

Matthews, B.W. Solvent content of protein crystals. *J. Mol. Biol.* 33, 491–497 (1968).

McClelland, S.E., Dryden, D.T.F., and Szczelkun, M.D. Continuous assays for DNA translocation using fluorescent triplex dissociation: application to type I restriction endonucleases. *J. Mol. Biol.* 348, 895–915 (2005).

McClell, J.L., and Rumelhart, D.E. *in* Explorations in parallel distributed processing. MIT Press, Cambridge, MA pp. 318-362 (1988).

McClelland, S.E., and Szczelkun, M.D. *Restriction Endonucleases, Nucleic Acids and Molecular Biology*. Germany: Springer Verlag; 14, 111–135 (2004).

Mernagh, D.R., Janscak, P., Firman, K. and Kneale, G.G. Protein-protein and protein-DNA interactions in the type I restriction endonuclease R.EcoR124I. *Biol. Chem.* 379, 497–503 (1998).

Mernagh, D.R., and Kneale, G.G. High resolution footprinting of a type I methyltransferase reveals a large structural distortion within the DNA recognition site. *Nucleic Acids Res.* 24(24), 4853-4858 (1996).

Mernagh, D.R., Reynolds, L.A., and Kneale, G.G. DNA binding and subunit interactions in the type I methyltransferase M.EcoR124I. *Nucleic Acids Res.* 25(5), 987-991 (1997).

Meselson, M., and Yuan. R. DNA restriction enzyme from *E. coli*. *Nature.* 217, 1110–1114 (1968).

Murray, N.E. Type I restriction systems: Sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* 64, 412–434 (2000).

Murray, N.E., Batten, P.L., and Murray, K. Restriction of bacteriophage λ by *Escherichia coli* K. *J. Mol. Biol.* 81, 395-407 (1973).

Murray, N.E., Daniel, A.S., Cowan, G.M., and Sharp, P.M. Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol. Microbiol.* 9(1), 133-143 (1993).

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* D53, 240-255 (1997).

Niv, M.Y., Ripoll, D.R., Vila, J.A., Liwo, A., Vanamee, E.S., Aggarwal, A.K., Weinstein, H., and Scheraga, H.A. Topology of type II REases revisited; structural classes and the common conserved core. *Nucl. Acids Res.* 35, 2227-2237 (2007).

Obarska, A., Blundell, A., Feder, M., Vejsadova, S., Sisakova, E., Weiserova, M., Bujnicki, J.M. and Firman, K. Structural model for the multisubunit type Ic restriction-modification DNA methyltransferase M.EcoR124I in complex with DNA. *Nucl. Acids Res.* 34, 1992-2005 (2006).

Obarska-Kosinska, A., Taylor, J.E., Callow, P., Orłowski, J., Bujnicki, J.M., Kneale, G.G. HsdR subunit of the type I restriction-modification enzyme EcoR124I: biophysical characterisation and structural modelling. *J. Mol. Biol.* 376(2), 438-452 (2008).

Otwinowski, Z., and Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* 276, 307–326 (1997).

Painter, J., and Merritt, E.A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *J. Appl. Cryst.* 39, 109-111 (2006).

Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S., and Tucker, P.A. Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D Biol. Crystallogr.* 61, 449-457 (2005).

Perrakis, A., Morris, R.J., and Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* 6, 458-463 (1999).

Powell, L.M., Dryden, D.T.F. and Murray, N.E. Sequence-specific DNA binding by Eco KI, a type Ia DNA restriction enzyme. *J. Mol. Biol.* 283, 963-976 (1998).

Powell, L.M., Dryden, D.T.F., Willcock, D.F., Pain, R.H. and Murray, N.E. DNA Recognition by the EcoK methyltransferase. *J. Mol. Biol.* 234, 60-71 (1993).

Price, C., Shepherd, J.C., Bickle, T.A. DNA recognition by a new family of type I restriction enzymes: a unique relationship between two different DNA specificities. *EMBO J.* 6(5), 1493-1497 (1987).

Rhodes, G. "Crystallography Made Crystal Clear" Academic Press, 1250 Sixth Ave., San Diego, CA 92101-4311 (2006).

Roberts, et al. The standard nomenclature for restriction enzymes, DNA methyltransferases and related proteins can be found in *Nucl. Acids Res.* 31, 1805-1812 (2005).

Rocak, S., and Linder, P. DEAD-box proteins: the driving forces behind RNA metabolism, *Nat. Rev. Mol. Cell Biol.* 5, 232-241 (2004),

Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J. et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucl. Acids Res.* 31, 1805-1812 (2003).

Rost, B., and Sander, C. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 19, 55-72 (1994).

Rubinson, K.A. Ladner, J.E. Tordova M. and Gilliland G.L. Cryosalts: suppression of ice formation in macromolecular crystallography *Acta Cryst.* D56, 996-1001 (2000).

Schneider, T.R., and Sheldrick, G.M. Substructure solution with SHELXD. *Acta Crystallogr.* D58, 1772-1779 (2004).

Schuettelkopf, A.W., and van Aalten, D.M.F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.* D60, 1355-1363 (2004).

Seidel, R., Bloom, J.G., Dekker, C., and Szczelkun, M.D. Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. *EMBO J.* 27(9), 1388-1398 (2008).

Seidel, R., Bloom, J.G., van Noort, J., Dutta, C.F., Dekker, N.H., Firman, K., Szczelkun, M.D., and Dekker, C. Dynamics of initiation, termination and reinitiation of DNA translocation by the motor protein EcoR124I. *EMBO J.* 24(23), 4188-4197 (2005).

Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. and Yokoyama, S. Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* 125, 287-300 (2006).

Szczelkun, M.D, and Dekker, C. Dynamics of initiation, termination and reinitiation of DNA translocation by the motor protein EcoR124I. *EMBO J.* 24(23), 4188-4197 (2005).

Seidel, R., van Noort, J., van der Scheer, C., Bloom, J.G., Dekker, N.H., Dutta, C.F., Blundell, A., Robinson, T., Firman, K. and Dekker, C. Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I. *Nat. Struct. Mol. Biol.* 11, 838–843 (2004).

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. Structure and mechanism of helicases and nucleic acid translocases. *Ann. Rev. Biochem.* 76, 23-50 (2007).

Singleton, M.R., Scaife, S. and Wigley, D.B. Structural analysis of DNA replication fork reversal by RecG. *Cell* 107, 79–89 (2001).

Singleton, M.R., and Wigley, D.B. Modularity and specialization in superfamily 1 and 2 helicases. *J. Bacteriol.* 184, 1819–1826 (2002).

Sistla, S., and Rao, D.N. S-Adenosyl-L-methionine-dependent restriction enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39(1), 1-19 (2004).

Stanley, L.K., Seidel, R., van der Scheer, C., Dekker, N.H., and Szczelkun, M.D., and Dekker, C. When a helicase is not a helicase: dsDNA tracking by the motor protein EcoR124I. *EMBO J.* 25(10), 2230-2239 (2006).

Stanley, L.K, and Szczelkun, M.D. Direct and random routing of a molecular motor protein at a DNA junction. *Nucleic Acids Res.* 34(16), 4387-4394 (2006).

Studier, F.W. and Bandyopadhyay, P.K. Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl Acad. Sci. USA.* 85, 4677–4681 (1988).

Suri, B., and Bickle, T.A. EcoA: the first member of a new family of type I restriction modification systems. Gene organization and enzymatic activities. *J. Mol. Biol.* 186(1), 77-85 (1985).

Suri, B., Shepherd, J.C.W. and Bickle, T.A. The EcoA restriction and modification system of *Escherichia coli* 15T-: enzyme structure and DNA recognition sequence. Cold Spring Harbor Symp. *Quant. Biol.* 43, 1217-1221 (1984).

Szczelkun, M.D., Dillingham, M.S., Janscak, P., Firman, K. and Halford, S.E. Repercussions of DNA tracking by the type Ic restriction endonuclease EcoR124I on linear, circular and catenated substrates. *EMBO J.* 15, 6335–6347 (1996a).

Szczelkun, M.D., Janscak, P., Firman, K., and Halford, S.E. Selection of non specific DNA cleavage sites by the type IC restriction endonuclease EcoR124I. *J. Mol. Biol.* 271(1), 112-123 (1997).

Taylor, I., Watts, D., and Kneale, G. Substrate recognition and selectivity in the type IC DNA modification methylase M.EcoR124I. *Nucleic Acids Res.* 21(21), 4929-4935 (1993).

Taylor, I.A., Davis, K.G., Watts, D., and Kneale, G.G. DNA-binding induces a major structural transition in a type I methyltransferase. *EMBO J.* 13(23), 5772-5778 (1994).

Taylor, I.A., Webb, M., and Kneale, G.G. Surface labelling of the type I methyltransferase M.EcoR124I reveals lysine residues critical for DNA binding. *J. Mol. Biol.* 258(1), 62-73 (1996).

Taylor, I., Patel, J., Firman, K. and Kneale, G.G. Purification and biochemical characterisation of the EcoR124 type I modification methylase. *Nucleic Acids Res.* 20, 179-186 (1992).

Theis, K., Chen, P.J., Skorvaga, M., Van Houten, B., and Kisker, C. Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *EMBO J.* 18, 6899-6907 (1999).

Titheradge, A.J.B., Ternent, D. and Murray, N.E. A third family of allelic hsd genes in *Salmonella enterica*: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA. *Mol. Microbiol.* 22, 437-447 (1996).

van Noort, J., van der Heijden, T., Dutta, C.F., Firman, K., and Dekker, C. Initiation of translocation by Type I restriction-modification enzymes is associated with a short DNA extrusion. *Nucleic Acids Res.* 32(22), 6540-6547 (2004).

Velankar, S.S., Soutanas, P., Dillingham, M.S., Subramanya, H.S. and Wigley, D.B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* 97, 75–84 (1999).

Wang, J., Cieplak, P., and Kollman, P.A. How well does a Restrained Electrostatic Potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.* 21, 1049-1074 (2000).

Webb, J.L., King, G., Ternent, D., Titheradge, A.J., and Murray, N.E. Restriction by EcoKI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J.* 15, 2003–2009 (1996).

Webb, M., Taylor, I.A., Firman, K., and Kneale, G.G. Probing the domain structure of the type IC DNA methyltransferase M.EcoR124I by limited proteolysis. *J. Mol. Biol.* 250(2), 181-190 (1995).

Weiserova, M., Dutta, C.F., and Firman, K.A. Novel mutant of the type I restriction-modification enzyme EcoR124I is altered at a key stage of the subunit assembly pathway. *J. Mol. Biol.* 304(3), 301-310 (2000).

Weiserova, M., and Firman, K. Isolation of a non-classical mutant of the DNA recognition subunit of the type I restriction endonuclease R.EcoR124I. *Biol. Chem.* 379 (4-5), 585-589 (1998).

Weiserova, M., Janscak, P., Benada, O., Hubacek, J., Zinkevich, V., Glover, S. W. and Firman, K. Cloning, production and characterisation of wild type and mutant forms of the R.EcoK endonucleases. *Nucl. Acids Res.* 21, 373-379 (1993).

Willcock, D.F., Dryden, D.T.F., and Murray, N.E. A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J.* 13, 3902-3908 (1994).

Winn, M., Isupov, M., and Murshudov, G. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr.* D57, 122-133. York, Cold Spring Harbor Laboratory Press (2001).

Yuan, R. Structure and mechanism of multifunctional restriction endonucleases. *Ann. Rev. Biochem.* 50, 285-315 (1981).

Yuan, R., Heywood, J. and Meselson, M. ATP hydrolysis by restriction endonuclease from *E. coli* K. *Nature New Biol.* 240, 42-43 (1972).

Yuan, R., Hamilton, D.L. and Burckhardt, J. DNA translocation by the restriction enzyme from *E. coli* K. *Cell* 20(1), 237-244. (1980).

Zinkevich, V., Popova, L., Kryukov, V., Abadjieva, A., Bogdarina, I., Janscak, P., and Firman, K. The HsdR subunit of R.EcoR124II: cloning and over-expression of the gene and unexpected properties of the subunit. *Nucleic Acids Res.* 25(3), 503-511 (1997).

APPENDIX

List of Presentations

Ettrich R., Csefalvai E., Kuta Smatanová I., Carey J, **Lapkouski M**: A translocation model for type I restriction-modification systems. Perspectives of Systems biology and ecology. Symposium. Nove Hradky. Czech Republic, 2008 (poster and lecture by R.Ettrich).

Mikalai Lapkouski, Panjikar Santosh, Pavel Janscak, Ivana Kuta Smatanova, Jannette Carey, Rüdiger Ettrich, Eva Csefalvai: Structure of the motor subunit and translocation model for EcoR124I restriction-modification complex. 22nd Annual Symposium of The Protein Society, San Diego, USA, 2008 (poster and lecture, given by Rudiger Ettrich).

Mikalai Lapkouski, Santosh Panjikar, Ivana Kuta Smatanova and Eva Csefalvai: Structural studies on the 120 kDa motor subunit (HsdR) of the EcoR124I endonuclease from *E. coli*. Keystone Symposia, Frontiers of Structural Biology. Sheraton Steamboat Resort, Steamboat Springs, Colorado, USA, 2008 (poster).

Alena Stsiapanava, Tana Koudelakova, **Mikalai Lapkouski**, Martina Pavlova, Jiri Damborsky and Ivana Kuta Smatanova: Crystallization and preliminary diffraction analysis of Rhodococcus rhodochrous NCIMB 13064 DhaA mutants. ICCBM12 12th International Conference on the Crystallization of Biological Macromolecules, Cancun, Mexico, 2008 (poster, presented by A. Stsiapanava).

Mikalai Lapkouski, Santosh Panjikar, Ivana Kuta Smatanova and Eva Csefalvai: Structural studies on the 120 kDa motor subunit (HsdR) of the EcoR124I endonuclease from *E. coli*. Herbsttagung der Gesellschaft für Biochemie und Molekularbiologie. Hamburg, Germany, 2007 (poster).

Mikalai Lapkouski, Santosh Panjekar, Ivana Kuta Smatanova and Eva Csefalvay: Structure of the 115 kDa DEAD-box motor subunit (HsdR) of the EcoR124I nuclease from *E. coli*. “The role of structures in biology – past, present and future” and a 3rd BIOXHIT Annual Meeting, Diamond, Didcot, UK, 2007 (poster).

Mikalai Lapkouski, Santosh Panjekar, Ivana Kuta Smatanova and Eva Csefalvay: Structural studies on the 120 kDa motor subunit (HsdR) of the EcoR124I endonuclease from *E. coli*. EMBO'07 - Exploiting Anomalous Scattering in Macromolecular Structure Determination, Grenoble, France, 2007 (poster).

Mikalai Lapkouski, Santosh Panjekar, Ivana Kuta Smatanova and Eva Csefalvay: Structure of the 115 kDa DEAD-box motor subunit (HsdR) of the EcoR124I nuclease from *E. coli*. 6th Discussions in Structural Molecular Biology and Bioinformatics. Mat. Struct. vol. 14 (1). Nove Hradky, Czech Republic, 2007 (poster).

Mikalai Lapkouski, Ivana Kuta Smatanova and Eva Jindrova: Overexpression, purification and crystallization attempts on the recombinant subunit EcoAI 5th Discussions in Structural Molecular Biology and Bioinformatics. Nove Hradky, Czech Republic, 2006 (poster).

List of Publications

1. **M. Lapkouski**, S. Panjikar, P. Janscak, IK. Smatanova, J. Carey, R. Ettrich, E. Csefalvay. Structure of the motor subunit and translocation model for EcoR124I restriction-modification complex. Accepted-in-principle to *Nature Str. Mol. Biol.* (2008).

ABSTRACT: EcoR124I is a multicomplex enzyme belonging to the Type I restriction-modification system from *E. coli*. Although, EcoR124I has been extensively characterized biochemically, there is no direct structural information about particular subunits. HsdR is a motor subunit that is responsible for ATP hydrolysis, DNA translocation and cleavage of the DNA substrate recognized by the complex. HsdR subunit was crystallized using a sitting-drop vapor-diffusion method. Crystals belong to the primitive monoclinic space group, with unit cell parameters $a = 85.75 \text{ \AA}$, $b = 124.71 \text{ \AA}$, $c = 128.37 \text{ \AA}$, $\beta = 108.14^\circ$. Native data were collected to 2.6 \AA resolution at the X12 synchrotron beamline of the EMBL Hamburg (Germany).

2. **Lapkouski, M.**, Panjikar, S., Kuta Smatanova, I., Csefalvay, E. Purification, crystallization and preliminary X-ray analysis of the HsdR subunit of the EcoR124I endonuclease from *Escherichia coli*, *Acta Cryst.* **F63**, 582-585 (2007).

ABSTRACT: Type I restriction-modification enzymes (RMs) act as conventional adenine methylases on hemi-methylated DNAs, but unmethylated recognition targets induce them to translocate thousands of basepairs before cleaving distant sites nonspecifically. The first crystal structure of a type I motor subunit responsible for translocation and cleavage suggests how the pentameric translocating complex is assembled and provides a structural framework for duplex DNA translocation by RecA-like ATPase motors.

3. A. Stsiapanava, T. Koudelakova, **M. Lapkouski**, M. Pavlova, J. Damborsky and I. Kuta Smatanova. Crystals of *Rhodococcus rhodochrous* NCIMB 13064 DhaA mutants diffracted to ultra high resolution: crystallization and preliminary diffraction analysis. *Acta Cryst.* **F64**, 137-140 (2008).

ABSTRACT: The enzyme DhaA from *Rhodococcus rhodochrous* NCIMB 13064 belongs to the haloalkane dehalogenases, which catalyze the hydrolysis of haloalkanes to the corresponding alcohols. The haloalkane dehalogenase DhaA and its variants can be used to detoxify the industrial pollutant 1,2,3-trichloropropane (TCP). Three mutants named DhaA04, DhaA14 and DhaA15 were constructed in order to study the importance of tunnels connecting the buried active site with the surrounding solvent to the enzymatic activity. All protein mutants were crystallized using the sitting-drop vapour-diffusion method. The crystals of DhaA04 belonged to the orthorhombic space group P2₁2₁2₁, while the crystals of the other two mutants DhaA14 and DhaA15 belonged to the triclinic space group P1. Native data sets were collected for the DhaA04, DhaA14 and DhaA15 mutants at beamline X11 of EMBL, DESY, Hamburg to the high resolutions of 1.30, 0.95 and 1.15 Å, respectively.

4. J. Kohoutova, I. K. Smatanova, J. Brynda, **M. Lapkouski**, J.L. Revuelta, J.B. Arellano, R. Eittrich. Crystallization and preliminary crystallographic characterization of the PsbP protein from oxygen-evolving complex of photosystem II from *Spinacia oleracea*. Manuscript submitted to *Acta Crystallogr. F Struct. Biol. Cryst. Commun.* (2008).

ABSTRACT: The preliminary X-ray diffraction analysis of the extrinsic PsbP protein of photosystem II (PSII) from spinach (*Spinacia oleracea*) was determined using the N-terminal His-tagged recombinant protein PsbP overexpressed in *Escherichia coli*. The recombinant PsbP protein (the thrombin digested recombinant His-tagged PsbP) stored in Bis-tris buffer pH 6.00 was crystallized by sitting drop vapor diffusion technique using PEG 550 MME as a precipitant and zinc sulfate as an additive. SDS – PAGE of the dissolved crystal showed that crystals did not contain degradation products of recombinant PSP protein. PsbP crystals diffracted with a resolution of 2.06 Å in space group P2₁2₁2₁ with unit cell parameters a=38.68 Å, b=46.73 Å, c=88.9 Å.

Curriculum Vitae

Name Lapkouski Mikalai

Business address Czech Republic
Laboratory of Molecular Biology
Institute of Physical Biology,
Department of Structure and
Function of Proteins,
University of South Bohemia,
Zamek 136, 37333,
Nove Hradý



Place and date of birth 22 December 1982, Minsk, Belarus

Telephone +420 776582929
E-mail lapkouski@greentech.cz, gort105@tut.by

Education

- 2005- now** - Laboratory of Molecular Biology, Department of Structure and Function of Proteins, University of South Bohemia, Institute of Physical Biology, Nove Hradý, Czech Republic
PhD. Student, R. Ettrichs group. Supervisor Ing. E. Csefalvay, Ph.D
- 2000-2005** - Department of Biochemistry, Belarusian State University and Institute of Bioorganic Chemistry, Natl. Acad. Sci. Minsk, Belarus. Diploma degree in Biology and Chemistry. Graduated with honors.

Bachelor Research topic: “*Structural and functional characteristics of the recombinant rat membrane NADPH-cytochrome P450 reductase*” at the Institute of Bioorganic Chemistry, National Academy of Sciences of Belarus and Department of Biochemistry, Minsk, Belarus.

Diploma Thesis: “*Cloning, expression and purification of recombinant flavoproteins from prokaryotes and eukaryotes*” at the Institute of Bioorganic Chemistry, National Academy of Sciences of Belarus and Department of Biochemistry, Minsk, Belarus

Research Techniques

Molecular Biology: Protein expression, purification and characterization; spectroscopy techniques. Protein labeling for NMR and SAD/MAD selenomethionine structure determination.

Structural Biology: Protein crystallization, X-ray data collection on synchrotron facilities (EMBL-Hamburg; ESRF, Grenoble; APS, Argonne, Chicago), data integration and scaling (HKL-2000, MOSFLM)

Macromolecular structure determination using SAD/MAD (SHELX(D,E)), structure refinement (CNS, REFMAC5, SHELXL), model building (COOT), structure analysis.

Research Cooperation . August, November, December 2006, February 2007 – visiting student at EMBL-Hamburg, Unit of structural biology. (to Santosh Panjikar, PhD).

Awards . 1) EMBO Short-Term Fellowship, 1 August – 1 November 2007 at EMBL- Hamburg, DESY
2) Keystone Symposium Fellowship, to visit a Keystone Symposia in Steamboat Springs, USA
3) Finn Wold Travel Award to visit 22nd Annual Symposium of The Protein Society in San Diego, USA.

Structures in PDB - 2RG1, 2V9Z, 2W00, 2VU4 (structures of DhaA4, DhaA14, DhaA15 are in preparation).

Courses and conferences attended

2008 - 22nd Annual Symposium of The Protein Society, San Diego, CA, USA

2008 - CCP4 school: From data processing to structure refinement and beyond. 23-28/05/2008 APS, Argonne Laboratory, Chicago, IL, USA

2008 - Keystone Symposia, Frontiers of Structural Biology. Sheraton Steamboat Resort Steamboat Springs, CO, USA

2007 - “Training in methods for Macromolecular Crystallography” - M2M-7: From Measurement to Model, November at EMBL/DESY, Hamburg.

2007 - “The role of structures in biology – past, present and future” and a 3rd BIOXHIT Annual Meeting, Diamond, Didcot, UK

2007 – “EMBO'07 - Exploiting Anomalous Scattering in Macromolecular Structure Determination”, Grenoble, France.

2007 – “Crystallization of biological macromolecules”, Nove Hradky, Czech Republic