UNIVERSITY OF SOUTH BOHEMIA IN ČESKÉ BUDĚJOVICE
FACULTY OF SCIENCE

Structure and Function of the C-terminal Domain of the HsdR Subunit from the Type I Restriction-Modification System EcoR124

PhD THESIS

Pavel Grinkevich, MSc

Supervisor: Prof. Rüdiger Ettrich, PhD
University of South Bohemia, České Budějovice
Faculty of Science, Institute of Chemistry
& Center for Nanobiology and Structural Biology
Institute of Microbiology,
Academy of Sciences of the Czech Republic

České Budějovice 2018

This thesis should be cited as:

Grinkevich P., 2018: Structure and Function of the C-terminal Domain of the HsdR Subunit from the Type I Restriction-Modification System EcoR124. PhD Thesis Series, No. 10, University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic, 119 pp.

Annotation

The Type I restriction-modification enzyme EcoR124 is a pentameric complex consisting of one specificity subunit, two methylation subunits and two motor subunits (HsdR) that can recognize specific DNA sequences and perform double-stranded DNA cleavage and modification. The HsdR subunit is responsible for ATP-dependent DNA translocation and DNA cleavage. Even though the first crystal structure of HsdR was obtained ten years ago, a large part of the C-terminus has not been resolved in any HsdR structures to date. This dissertation aims to elucidate its role within the HsdR subunit and the whole pentameric complex by solving the structure of the C-terminus by means of X-ray diffraction crystallography and explore its function using biochemical, microbiological, bioinformatical and computational methods.

Declaration [in Czech]

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 úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, 10.09.2018 .....

Pavel Grinkevich

This thesis originated from a partnership of the Faculty of Science, University of South Bohemia, and the Center for Nanobiology and Structural Biology, Institute of Microbiology of the Czech Academy of Sciences, supporting doctoral studies in the Biophysics study program.

Financial support

Czech Science Foundation (P207/12/2323; http://www.gacr.cz/) and Czech Research Infrastructure for Systems Biology C4SYS (LM2015055).

The thesis is based on the following publications:

I. *Grinkevich P.*, Iermak I., Luedtke N.A., Mesters J.R., Ettrich R., Ludwig J. (2016) pHluorin-assisted expression, purification, crystallization and X-ray diffraction data analysis of the C-terminal domain of the HsdR subunit of the Escherichia coli type I restriction-modification system EcoR124I. *Acta Cryst. F***72**, 672-676 (IF_{2017} = 0.989).

PG devised the experiments (jointly with *RE* and *JL*), performed purification, and crystallization experiments (jointly with *NL*), collected and analyzed the X-ray diffraction data (jointly with *II* and *JM*), wrote the manuscript (jointly with *RE* and *JL*).

II. *Grinkevich P.*, Sinha D., Iermak I., Guzanova A., Weiserova M., Ludwig J., Mesters J.R., Ettrich, R.H. (2018) Crystal structure of a novel domain of the motor subunit of the Type I restriction enzyme EcoR124 involved in complex assembly and DNA binding. *J Biol Chem*, doi: 10.1074/jbc.RA118.003978. (IF_{2016} = 4.125).

PG devised the experiments (jointly with *RHE*), determined the structure (jointly with *JRM* and *II*), performed in vitro and in vivo experiments (jointly with *AG* and *MW*), performed structural (jointly with *DS* and *RHE*) and bioinformatical (jointly with *RHE*) analysis, wrote the manuscript (jointly with *RHE, JRM, JL, DS* and *II*).

III. Sinha D., *Grinkevich P.*, Ettrich, R.H. (2018) Functional characterization of the fifth domain in HsdR subunit of Type I RM system EcoRI24I: A molecular dynamics study. *Manuscript.*

PG performed structure determination and re-refinement (jointly with *DS* and *RH*) and contributed to data analysis and writing and editing the manuscript (jointly with *DS* and *RH*).

Acknowledgements

This work would not have been possible without a constant input from my supervisor, Rüdiger Ettrich, whose ample knowledge in protein structure and keen observations allowed me to proceed with the project in the circumstances when I otherwise saw little hope.

I could not be more grateful to my dear friend and colleague, Iuliia Iermak, who introduced me to the world of protein crystallography and uplifted me every step of my PhD journey. Another great crystallographer, Jeroen Mesters, provided me with great scientific insights during my internship in the beautiful city of Lübeck and thereafter.

I am very thankful to Jost Ludwig for sharing his vast knowledge on cloning techniques and vector design and helping me out in every way possible. I could not have completed *in vivo* and EMSA experiments without my colleagues from Marie Weiserová lab in Prague where I spent a very productive six months.

My sincere gratitude extends to all my friends both in Nové Hrady and České Budějovice: Saurabh Kumar Pandey, Katsiaryna Shamayeva, Tatyana Prudnikova, Vitali Bialevich, Hanna Huryna and others. Lastly, I thank my family, my little sister Anya, my mom Yanina and my father Iosif, for their constant support during my PhD studies.

Dedicated to my family and friends

Contents

List of abbreviations

ADP	adenosine diphosphate
ATP	adenosine triphosphate
CR	central region (HsdS)
CTD	C-terminal domain
C-terminus	carboxyl-terminus
CTP	cytidine triphosphate
dATP	deoxyadenosine triphosphate
DSSP	Define Secondary Structure of Proteins (algorithm)
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
EOP	efficiency of plating
GTP	guanosine triphosphate
HGT	horizontal gene transfer
Hsd	host specificity determinant
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilobase
MD	molecular dynamics
MTase	methyltransferase
NTD	N-terminal domain
N-terminus	amino-terminus
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction

PDB	Protein Data Base
REase	restriction endonuclease
R-M	restriction-modification
SAM	S-Adenosyl methionine
SDS	sodium dodecyl sulfate
TRD	target recognition domain
UTP	uridine triphosphate
WT	wild type

1. INTRODUCTION

1.1. Prologue and aims of the research

DNA as the central molecule of hereditary information was undoubtedly one of the major discoveries of the 20th century biology. The double-helix model of DNA, based on one X-ray diffraction image taken by Rosalind Franklin and Raymond Gosling, was suggested by Francis Crick and James Watson and presented in a series of *Nature* articles in 1953. Realizing the key role of DNA as information storage molecule in every living cell brought DNA maintenance machinery to the forefront of molecular biology research. While replication and transcription are the most important processes in DNA maintenance, the subtleties in expression and regulation of genetic material are seemingly endless.

DNA restriction-modification (R-M) enzymes, discovered in the same decade as structure of DNA, belong to a category of proteins that modulate maintenance of genetic material in different, often unexpected, ways. First observed as an outcome of bacteria fending off bacteriophage attacks, R-M enzymes were later shown to be able to distinguish between self and non-self DNA by reading out chemical modifications in DNA bases and cutting unmodified foreign DNA. Since R-M enzymes can recognize a wide variety of specific DNA sequences, they were likened to a prokaryotic immune system. The modern view of R-M enzymes ascribes a number of functions to them, such as regulation of gene expression, driving divergent evolution, a role in horizontal gene transfer and pathogenicity.

In the 1970s, isolation of R-M enzymes paved the way for modern techniques of recombinant DNA, allowing to study and manipulate individual genes from various organisms for the first time. These workhorses of molecular biology, later named Type II R-M enzymes, have been used in laboratories ever since, owning to their predictable DNA cutting behavior within or near a specific DNA recognition site. EcoR124 is a member of Type I R-M enzymes whose cleavage sites are, on the contrary, unpredictable. This behavior was later explained by the ability of Type I R-M enzymes to pull DNA molecules through themselves, a phenomenon called DNA translocation. While this feature did not allow EcoR124 to be used as a predictable DNA cutter in the laboratory, other proba-

ble uses, such as DNA-binding drug discovery and as a molecular motor in biosensors, remain a possibility (1). EcoR124 is also one of the better-studied Type I R-M enzymes and serves as a model for the whole class.

EcoR124 holoenzyme is a multimeric complex formed from three different subunits: HsdS (responsible for sequence-specific DNA recognition), HsdM (DNA methylation) and HsdR (ATP hydrolysis, DNA cutting and translocation). HsdR subunit is a large protein that consists of multiple domains and combines both endonuclease and translocation activities. Even though the first crystal structure of HsdR was solved almost ten years ago (2), a large part of the C-terminus (~150 residues) was not resolved in that and five subsequent structures.

The main research goal of this work was to solve the missing part of HsdR structure and elaborate on the possible function of the C-terminal domain within the HsdR subunit and the whole EcoR124 complex. The following steps were undertaken to achieve this goal:

1. Develop a strategy to express and purify the C-terminus of HsdR.
2. Work out and optimize crystallization procedure.
3. Collect and analyze X-ray diffraction data and solve the structure of the C-terminal domain.
4. Produce a relevant model of the full-length HsdR subunit combining the previously solved wild-type and the C-terminal domain structures.
5. Conduct biochemical and microbiological experiments to elucidate the biological role of the C-terminal domain.
6. Perform bioinformatical and structural evaluation of the C-terminal domain.

1.2. R-M enzymes

Restriction-modification (R-M) enzymes constitute a diverse polyphyletic group of DNA cutters and modifiers widespread in prokaryotic microorganisms. United in their ability to recognize specific DNA sequences, cut double-stranded DNA and modify DNA bases by methylation, R-M systems are part of prokaryotic immune system, differential gene expression, horizontal gene transfer and other important aspects of bacterial physiology and genetics. This chapter introduces four major groups of R-M enzymes and their representatives and highlights their role in biology and evolution of prokaryotes. The historical sketch mostly focuses on Type I R-M enzymes.

1.2.1. Historical overview

The subjects of this dissertation, EcoR124I and the almost identical EcoR124II, belong to Type I R-M enzymes, incidentally the first group of R-M enzymes to have been discovered. Effects of the Type I R-M systems were described in the early 1950s, constituting the ability of certain bacterial strains to restrict or enlarge the host range of bacteriophages after only one growth cycle. This phenomenon was called 'host-controlled variation in bacterial viruses' (3,4,5). Most importantly, this phenomenon could not be attributed to a mutation, since a single growth cycle was enough to revert the phenotype.

A decade later, host controlled variation in bacteriophages was revealed to be the result of DNA methylation and degradation occurring in bacterial cell (6,7,8). It was learned that restriction (R) of bacteriophage host range was caused by the ability of the host cell to degrade the unmodified viral DNA. Modification (M), in contrast, widens host range, protecting viral DNA by methylation.

The year 1968 heralded independent isolation and biochemical characterization of the first two Type I R-M enzymes from *E. coli* strains K-12 and B (9,10), which later became known as EcoKI and EcoB. Even though the plasmid R124 carrying EcoR124 was identified as early as 1966 (11), the first isolation and biochemical description followed only eleven years later (12). Initially, purified EcoR124 samples were plagued by the presence of extraneous ATP-independent nuclease activity (1). Adding to the confusion, the plasmid's original source was mislabeled as *E. coli*, even though it came from *Salmonella enterica* serovar

TABLE 1

Known structures of proteins from Type I R-M family in chronological order.

Protein name	*Source*	*Resolution (Å)*	*PDB ID(s)*	*Year*	*Reference*	*Comments*
M.EcoKI	*E. coli*	2.8	2AR0	2005	-	-
HsdS	*Methanocaldococcus jannaschii*	2.4	1YF2	2005	(13)	-
HsdS	*Mycoplasma genitalium*	2.3	1YDX	2005	(14)	
HsdM	*Bacteroides thetaiotaomicron*	2.2	2OKC	2007	-	
R.EcoR124	*E. coli*	2.6	2W00	2008	(2)	
HsdR	*Vibrio vulnificus*	2.3	3H1T	2009	(15)	Fragment
HsdM	*Methanosarchina mazei*	2.55	3KHK	2009	-	-
MTase.EcoKI	*E. coli*	18	2Y7C, 2Y7H	2009	(16)	Ocr-bound and free
HsdM	*Streptococcus thermophilus*	2.25	3LKD	2010	-	-
HsdS	*Thermoanaerobacter tengcongensis*	1.95	3OKG	2011	(17)	-
HsdM	*V. vulnificus*	1.8	3UFB	2012	(18)	
R.EcoR124	*E. coli*	2.74-2.99	4BE7, 4BEB, 4BEC	2014	(19)	K220R, K220E and K220A mutants
R.EcoR124	*E. coli*	2.4	4XJX	2016	-	E165H mutant
MTase	*T. tengcongensis*	3.2	5YBB	2017	(20)	DNA bound
R.EcoR124	*E. coli*	2.45	5J3N	2017	*Paper II*	C-terminal domain fused to pHluorin
R.EcoR124	*E. coli*	2.6	6H2J	2018	*Paper II*	Re-refined full-length structure

Typhimurium; the right name for EcoR124 should therefore have been StyR124 (1).

Even though the first X-ray crystal structure of a protein was solved in late 1950s (21), the structures of the individual subunits from Type I R-M family only began to emerge in 2000s. The structures of the first HsdM and HsdS subunits (13) appeared in 2005, while the first HsdR structure was solved in our group in 2009 and came from EcoR124 (2). A few other structures of wild-type and mutant subunits have been added to PDB since, along with a EM single particle model of a methyltransferase (MTase) (16) and a recent 3.2 Å DNA-bound MTase crystal structure (20). For further information on all structures of Type I R-M enzymes available to date, see Table 1.

As evidenced by previous decades of research, crystal structures of the Type I R-M enzymes have been elusive. Further research is somewhat hindered by a rather incomplete and chaotic nature of the structural information available: the subunits come from various source organisms, belong to different R-M systems and often lack crucial parts. The main aim of this dissertation is to complete the HsdR structure, and in doing so, it is our hope that a fuller picture of R-M systems, and specifically EcoR124, will emerge.

1.2.2. Diversity and classification

R-M enzymes is a very diverse group divided into four types, I–IV, based on their subunit organization, cofactor requirements and DNA cleavage character (22,23). Around 8000 restriction endonucleases (REases) and MTases have been biochemically or genetically described and hundreds are added to the list every year, while the number of putative enzymes has increased dramatically since the advent of whole genome sequencing (Figure 1). Although the emphasis of this work is a Type I R-M system, a brief description and distinctive features of the R-M types are given below and summarized in Table 2.

Type I R-M enzymes were first to be isolated and described as already pointed out in the previous section. These systems consist of three different kinds of subunits (HsdS, HsdM and HsdR, where hsd stands for ‘host specificity determinant’) that dynamically come together to form complexes performing various

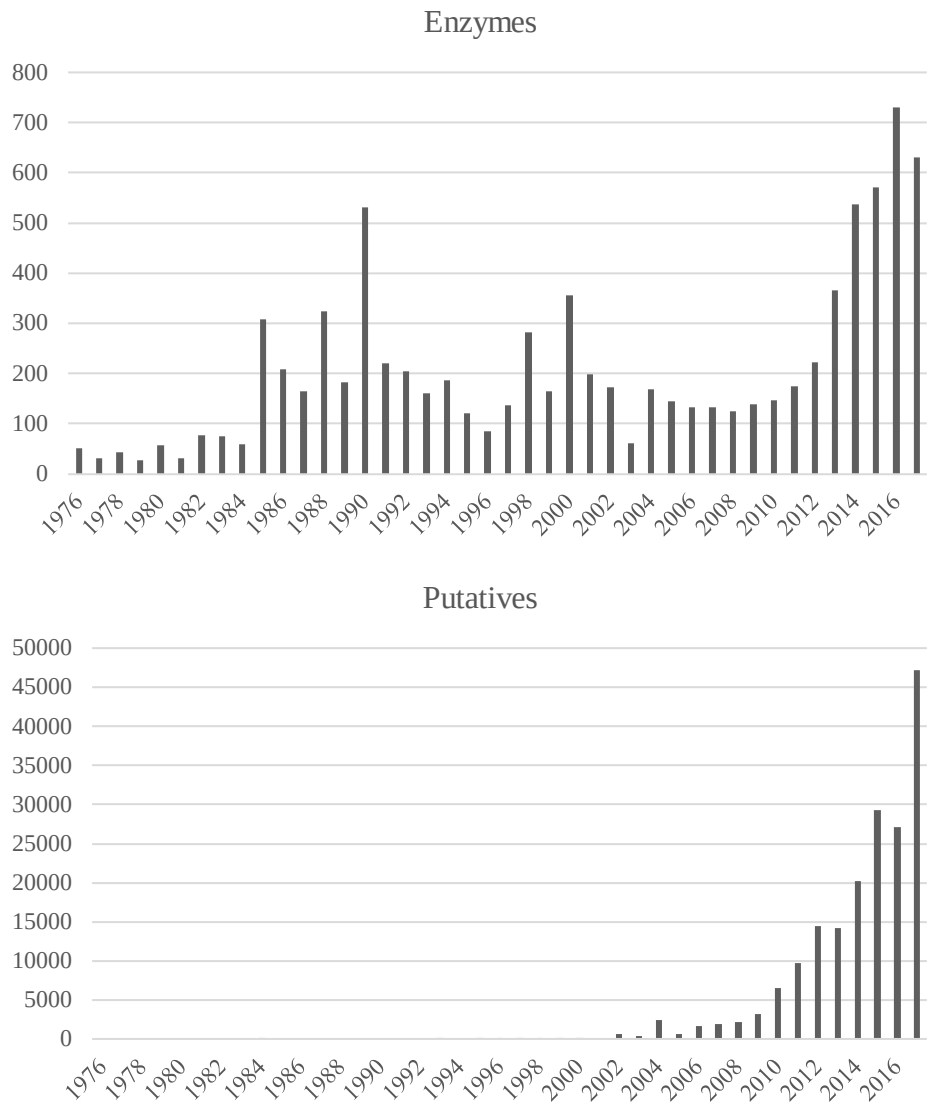

FIGURE 1

Total number of actual (on the top) and putative (on the bottom) R-M enzymes added to REBASE per year from 1976 until 2017. Note that HsdR, HsdM and HsdS subunits are counted as separate enzymes. The information was retrieved from http://rebase.neb.com/rebase/rebenzyearbar.html on Jan 11, 2018.

tasks. Most notably, one HsdS and two HsdM subunits form a trimeric methyltransferase (MTase) that is able to recognize a specific DNA sequence and trans-

fer a methyl group to an adenine (normally) within this sequence. The recognition capability is harbored in HsdS (or simply S for 'specificity'), while methyl transfer is done by HsdM (or M for 'modification'). MTase can further recruit two HsdR subunits (R for 'restriction') to form the full pentameric complex. This complex is capable of ATP-dependent DNA translocation and a subsequent DNA cleavage at a distant random site (24). Both activities are only exhibited by the whole complex but are catalyzed by HsdR. The recognition sequence consists of two different parts ('bipartite') separated by an unspecific spacer of 5-8 nucleotides (25). This peculiar recognition sequence design corresponds to the HsdS structure that consists of two globular target recognition domains (TRDs) connected by helical dimerization domains. Type I R-M enzymes are further divided into 5 families, A through E, based on antibodies cross-reactivity and DNA hybridization studies. Typical examples include EcoKI (Type IA), EcoAI (Type IB) and EcoR124I (Type IC). The features of Type I R-M enzymes are summarized in Table 2. For a recent review, see Loenen *et al.* (25).

TABLE 2

Types of R-M systems and their features.

	Type I	*Type II*	*Type III*	*Type IV*
Subunits	HsdR, HsdS, HsdM, (R, S, M)	R, M	Res, Mod (R, M)	Various composition (examples given in text)
MTase	M_2S	M	M_2	
REase	R_2M_2S	R_2 (usually)	R_2M_2	
Cofactors	SAM, ATP, Mg^{2+}	Mg^{2+} (usually)	SAM, ATP, Mg^{2+}	GTP or other NTPs
Recognition sites	Bipartite, asymmetric	Palindromic (most often) or asymmetric	Nonpalindromic, 2 required for cleavage	Various, containing modified nucleotides
Cleavage site	Random, distant	Within or next to recognition site	~25–27 bp from recognition site	Within or next to recognition site
REBASE[1]	115 enzymes, 17542 putative	4189 enzymes, 26962 putative	22 enzymes, 7042 putative	19 enzymes, 10870 putative
Examples	EcoKI, EcoR124, EcoAI	EcoRI, EcoRV, HindIII	EcoP1, EcoP15, HinfIII	MrcA, MrcBC, Mrr

[1] retrieved from http://rebase.neb.com/cgi-bin/statlist on Jan 11, 2018.

Type II R-M enzymes possess a simpler structural organization and have been frequently referred to as the workhorses of molecular biology due to their regular use as sequence-specific DNA cutters in laboratories. Consisting of two independently-acting proteins, the restriction endonuclease (REase) and methyltransferase (MTase), these R-M systems recognize DNA sequences specific to each enzyme and cleave DNA within or next to the recognition locus. REases belonging to Type II mainly act as homodimers and cleave both DNA strands simultaneously, but other variations exist, e.g., consequent cleavage of DNA strands by two monomers, dimerization of homodimers, transient dimers, etc. Almost all Type II REases require divalent cations, such as Mg^{2+}, as cofactors. Divided into at least 11 subtypes, EcoRI, EcoRV and HindIII are the most common representatives and the most well-studied (Table 2). Intricacies of Type II R-M enzymes are explained in a recent review (26).

Type III R-M enzymes appear to be present in a large number of bacteria even though only a few have been biochemically described (Table 2). The EcoP1, EcoP15 and HinfIII enzymes, the most studied to date, consist of the larger Res and smaller Mod subunits (27) that form the Res_2Mod_2 or an independent Mod_2 complexes. The enzymes require ATP, Mg^{2+} and S-Adenosyl methionine (SAM) to function (28,29). Two asymmetrical recognition sites in the opposite orientation are required for DNA cleavage to occur while even a single recognition site can be methylated (30). Type III R-M enzymes cleave DNA ~25–27 bp downstream the recognition site and produce 5′-overhangs of 2–3 nucleotides. Later experiments observed ATP-dependent translocation between the enzymes attached to two closest recognition sites in head-to-head orientation (31). Type III REases have been used in Serial Analysis of Gene Expression (SAGE) experiments for their ability to produce longer 26 bp tags beneficial for an efficient gene identification by BLAST searches. This and other advances in Type III R-M systems are discussed at length in a recent review (32).

Type IV R-M enzymes differ from the other three types in their methylation-dependent DNA cleavage. The better-described representatives include MrcA MrcBC and Mrr. GTP, ATP, dATP, UTP and CTP may be required for cleavage activity by various members (33). The structural complexity of Type IV R-M enzymes is exemplified by MrcBC encoded by the *mrcB* and *mrcC* genes. Two products are expressed from *mrcB*, the full-length $MrcB_L$ and a shorter $MrcB_S$.

Both products form heptameric rings in presence of GTP and its non-hydrolysable analogs. The heptameric structures can then bind MrcC. The $MrcB_LC$ complex is the functional enzyme that cleaves DNA upon complex dimerization via MrcC subunits preceded by DNA translocation. The $MrcB_SC$ complex appears to modulate the excess of free MrcC subunit, serving a regulatory purpose. The process is described in detail by Bourniquel (34). Owning to their ability to cleave DNA at a fixed position from the modification site, the MspJI family of Type IV R-M enzymes was suggested as a tool for epigenetic studies (35). Further information on Type IV R-M enzymes can be found in a recent review (33).

1.2.3. Evolutionary and biological role

The staggering diversity of R-M enzymes and their omnipresence in prokaryotic genomes prompts numerous questions about their biological significance. The protective function was hinted at during the initial experiments that led to the discovery of R-M enzymes (3,4,5). The ability to distinguish the self and non-self DNA makes R-M enzymes a part of bacterial immune system, protecting the host from bacteriophage invasion. Later on, further evidence began to emerge in support of other important roles these enzymes have in horizontal gene transfer, divergent evolution and virulence. These and other biological features of R-M systems are briefly discussed below.

The role of R-M systems as a defense mechanism from bacteriophages was postulated since their discovery. The classical bacteriophage assay shows that presence of an R-M system limits the unmodified phage propagation by factors of 10^4–10^8 for EcoKI (36), and 10^4 and 10^2 for EcoR124I and EcoAI, respectively (37). An environmental survey of restriction resistance of naturally occurring coliphages to the Type I and II R-M systems of *E. coli* showed a moderate success in using restriction as a defense mechanism. Of all tested phage–R-M system pairs, 26% were restriction-proficient for at least one Type I R-M system and 68% for one Type II (38). In all, 24 distinct phages were studied, along with three Type I and four Type II R-M enzymes. It was suggested that R-M systems could be useful for *E. coli* in colonizing new environments, greatly enhancing the likelihood of their becoming established; this hypothesis was tested by experiments and mathematical modeling (39). Based on a transient effect of R-M

systems as a defense mechanism, the same authors suggest that colonization selection may play a major role in R-M maintenance and evolution (39).

Horizontal gene transfer (HGT) is a major factor in evolution of prokaryotes. Prokaryotic populations often exchange genetic information, both intra- and interspecifically, which promotes a fast propagation of new traits among cohabitating microorganisms. While the importance of HGT in bacterial evolution is not disputed (40), the involvement of R-M systems in this process is still not fully understood. A study of 43 pan-genomes showed an increased number of R-M systems in naturally competent hosts, such as *Helicobacter pylori* (41). However, the authors refrained from concluding whether this over-representation conveys an advantage for naturally transformable bacteria or is simply the result of R-M systems being more frequently acquired during transformation.

Divergent evolution in bacteria as a result of decreased HGT between different strains constituting a species is sometimes also attributed to R-M systems. Strains with incompatible R-M systems become more isolated because of lower transformation rates due to non-self DNA degradation upon its acquisition. For example, a survey of 20 genomes of an important human pathogen *Neisseria meningitidis* showed that a unique combination of 22 R-M systems determined the polyphyletic clades the genomes were classified into (42). Among 22 putative R-M systems that were identified, 14 belonged to Type II, four to Type III and two to Type I. A similar correlation between R-M systems and phylogenetic clades was reported for another human pathogen *Haemophilus influenzae*, concluding that such peculiarities in R-M system distribution can limit genetic exchange between phylogenic groups (43). Another evolutionary pathway is particular to the Type I R-M enzymes, where a new specificity can be acquired via permutations of target recognition sequence in *hsdS* gene, owing to the fact that the HsdM and HsdR subunits are reused in both DNA methylation and restriction. A link between the changing specificity of Type I systems and an increase in genetic heterogeneity was also suggested in such important pathogens as *Streptococcus pneumoniae* (44), *Helicobacter pylori* (45) and *Mycoplasma pulmonis* (46). The above examples imply at least some degree of involvement of R-M enzymes in divergent evolution of bacteria, even though it is hard to predict the scope and universality of such involvement.

Regulation of gene expression via methylation is another biological role various R-M systems and stand-alone MTases have been shown to participate in. Rapid alternations in epigenome are especially efficient upon entering a new environment and are an example of phase variation. Methylation of specific sequences may lead to a massive shift in gene expression by modulating the binding of transcription factors. For example, the Type I R-M system SpnD39III is involved in phase switching of phenotypic forms in *Streptococcus pneumoniae* (47). In this case, genetic rearrangements in *HsdS* gene lead to six alternative specificities with defined methylation patterns, producing phenotypes with distinct virulence. A similar phase variation mechanism was suggested for *Helicobacter pylori*, where changing the specificity of HsdS through recombination of TRDs produces different methylation patterns throughout the genome resulting in changes in gene expression (48). These and other numerous examples of methylation-mediated regulation in expression patterns and phenotypic variation show a promising start to a better understanding of R-M systems importance in such processes as differential gene expression, virulence and immune evasion in bacteria [see (49) for a review].

Other functions have been ascribed to R-M enzymes, such as involvement in homologous recombination (50), stabilizing persistence of the plasmids that code for R-M systems through post-segregational killing (51), degradation of human chromosomal DNA and disruption of mitosis by a pathogenic bacterium (52), etc. It is quite possible that this list, by no means exhaustive, will continue to be appended due to an extreme structural variability and diversity of R-M systems. A deeper understanding of biology of R-M enzymes will help to explain such important questions as virulence of pathogenic bacteria, emergence and intra- and interspecific transmission of new traits including the medically relevant multiple drug resistance and others.

1.3. EcoR124 as a model Type IC R-M enzyme

As pointed out in the previous sections, the first R-M enzymes to be observed and discovered were later classified as Type I. Together with EcoK and EcoB, EcoR124 was among the pioneering few and remains an important model for the whole type to this date. This role as a model enzyme is attributed to the sheer amount of biochemical and structural insights that have been accumulated in half a century of research. This chapter deals with the discovery, genetics and structure of EcoR124, its subunits and their enzymatic activities.

1.3.1. Discovery and genetic organization

The name of EcoR124 already contains a hint of history in it: the plasmid in which the *hsd* genes were subsequently identified was first mentioned as R124 by Meynell and Datta in 1966 (11). The authors tested a number of drug resistance factors, or R factors (an old name for plasmids that convey drug resistance), for their antibiotic resistance and phage sensitivity. The R124 and other R factors originated from the strains of *Salmonella enterica* serovar Typhimurium sent to the Enteric Reference Laboratory in London, hence the later confusion about the source organism and the erroneous 'eco' (from *E. coli*) prefix in the name. In 1968, Bannister and Glover showed that R124 factor is able to restrict the growth of bacteriophages λ, P2 and φ80 as well as modify them (53).

Early research of EcoR124 mostly focused on genetical aspects of the plasmid R124 and later discovered R124/3 (12). Restriction fragment analysis produced a detailed map for R124 (54) and later studies identified approximate locations of *hsd* genes. An adapted version of R124 restriction map produced by the EcoR1 endonuclease and the locus of *hsd* genes is shown in Figure 2. Further genetical studies led to an understanding that the differences between R124 and R124/3 are due to DNA rearrangements that cause a change in enzyme specificity (55). These two plasmids express the allelic EcoR124 and EcoR124/3 enzymes (the modern equivalents are EcoR124I and EcoR124II; EcoR124 is used to refer to both forms), and they indeed differ in the sequence of their respective *hsdS* genes.

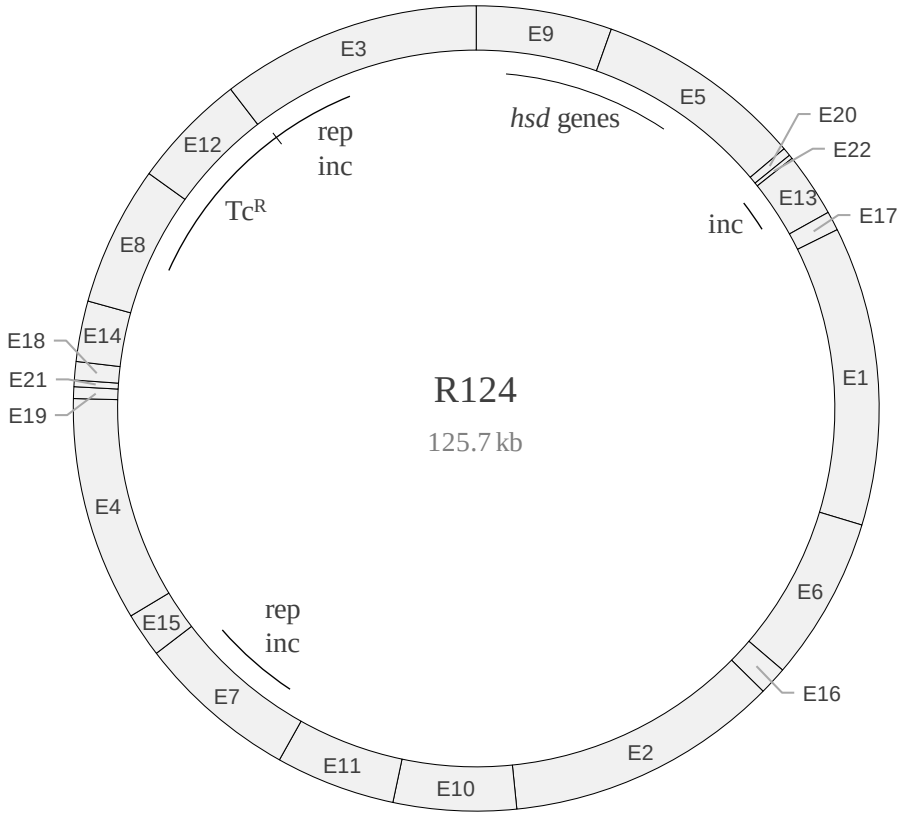

FIGURE 2

The restriction fragment map of the R124 plasmid produced by EcoRI; adopted from Youell and Firman (1) and based on the map produced by Campbell and Mee (54). Approximate loci are shown for *hsd* genes, tetracycline resistance gene (Tc^R), replication (rep) and incompatibility (inc) regions.

The complications with R124 size and stability prompted a shift to studying the enzymes themselves once *hsd* genes were cloned in 1980s (1). The difference between specificities of EcoR124I and EcoR124II was also established during that time and turned out to be simply an additional nucleotide in the spacer region of the recognition sequence (56). When *hsdM* and *hsdS* genes were sequenced, the allelic forms of EcoR124 were found to be identical except for an insertion of 12 bp in the *hsdS* of EcoR124II that is attributed to change in specificity (57). A mistake in *hsdR* sequence in that study was later corrected, chang-

FIGURE 3

Hsd region of EcoR124 adapted from Price *et al.* (57). The length of the genes is given according to rebase.neb.com. The size of *HsdR* and *HsdM* is identical for the allelic EcoR124I and EcoR124II, while *HsdS* of EcoR124II is 15 bp longer. *HsdM* and *hsdS* sequences overlap by 1 bp and form a single operon, while *hsdR* is transcribed from a separate promoter. The arrows denote the direction of transcription.

ing 12 C-terminal amino acids (1022-1034) and adding a further five amino acids, bringing the total length of HsdR to 1038 (http://rebase.neb.com/cgi-bin/seqget?EcoR124I). The *HsdR* sequence of the allelic forms of EcoR124 is identical.

The transcription order of *hsd* genes of EcoR124 is *hsdM-hsdS-hsdR*. *HsdM* and *hsdS* are expressed from a single operon while *hsdR* is transcribed from a separate promoter (Figure 3). Typical for co-translating genes, the start codon of *hsdS* sequence overlaps by one nucleotide with the stop codon of *hsdM*.

1.3.2. HsdS – the specificity subunit

HsdS subunit, where S stands for specificity, is responsible for DNA recognition. It only functions as a part of either the trimeric MTase or the pentameric complex and is insoluble on its own in case of EcoR124 (58). While no crystal structure has been solved for S.EcoR124 (a shorthand for the 'HsdS of EcoR124') to date, three crystal structures of putative HsdS subunits from *Mycoplasma genitalium* and two from thermophilic microorganisms were obtained (Table 1). All structures exhibit a largely similar fold despite significant differences in sequence (Figure 4A). A homology model was produced for S.EcoR124 based on the putative HsdS crystal structures and other available information (59). The structure was later refined further as a part of the pentameric complex by Kennaway *et al.*, a model derived from electron microscopy and small-angle scattering data (60).

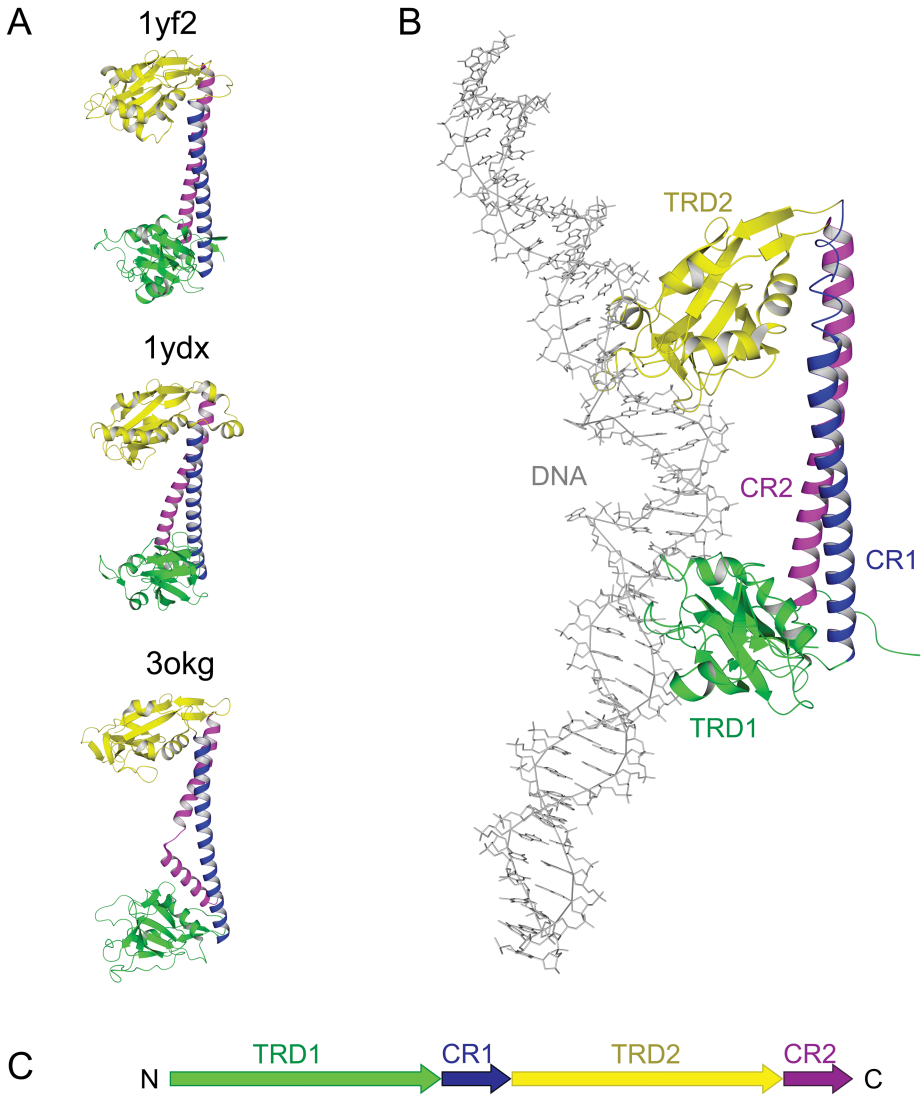

FIGURE 4

3D structure of HsdS subunits. (A) Known crystal structures of putative HsdS subunits from various sources: an archeon *Methanocaldococcus jannaschii* [PDB ID 1YF2, (13)], a human pathogen *Mycoplasma genitalium* [PDB ID 1YDX, (14)] and thermophilic bacterium *Thermoanaerobacter tengcongensis* [PDB ID 3OKG, (17)]. Homology model (B) and schematic representation (C) of HsdS with DNA (shown in gray) from EcoR124 reported by Obarska *et el.* [ftp://genesilico.pl/iamb/models/MTases/M.EcoR124I/M.S.EcoR124I.DNA.pdb, (59)]. Target recognition domains 1 and 2 (TRD1 and TRD2) and central regions 1 and 2 (CR1 and CR2) have the same coloring scheme in all structures. All subunits are shown in the same orientation after MUSTANG alignment (61) in YASARA (62).

HsdS subunit consists of two target recognition domains (TRDs) separated by central regions (CRs). TRDs are globular domains that specifically bind two parts of the bipartite DNA recognition site characteristic for Type I R-M enzymes, while CRs are comprised of two antiparallel α-helices bringing the TRDs to a distance corresponding to an unspecific spacer within the recognition sequence (Figure 4B). Sequence analysis showed that CR are more conserved, while TRDs exhibit more variation. Several Hybrid R-M enzymes with a novel specificity were created by combining TRDs from two different HsdS subunits. First shown for the StySB and StySP enzymes belonging to the Type IA family (63), hybrid R-M systems revealed that the left and right parts of a bipartite recognition sequence are recognized by TRD1 and TRD2, respectively (64). A similar hybrid R-M enzyme was later obtained for the EcoR124I and EcoDXXI, which prompted the authors to speculate that this reshuffling of specificities between allelic forms of Type I R-M enzymes could be an analog of prokaryotic immune system (65). This possibility is indeed viable due to the fact that in Type I R-M enzymes, a change of specificity is not as damaging for the host since the same HsdS subunit is involved in both methylation and restriction as a part of the trimeric MTase and pentameric complex, thus both specificities change simultaneously and do not lead to an unexpected degradation of host DNA.

CRs can also be involved in sequence specificity change, as illustrated by the allelic versions of HsdS subunits from EcoR124I and EcoR124II, both recognizing the same sequence 5′-GAA$(N)_x$RTCG-3′ where x is equal to 6 and 7 bp, respectively. An additional nucleotide in the gap region is accounted for with an extra 12 bp repeat in CR1, coding for Thr-Ala-Glu-Leu (TAEL). The TAEL repeat effectively adds an extra turn in the CR1 α-helix, bringing the TRDs of S.EcoR124II 0.34 nm further apart and rotating them 36° compared to EcoR124I (57).

Mutagenesis studies on S.EcoR124 carried out by Weiserova *et al.* (66,67) hinted at a possible interface between HsdS and HsdM. The Trp212→Arg mutation, lying in the CR1, was shown to significantly lower binding of HsdS to HsdM. DNA binding was also drastically reduced and only partially restored *in vitro* with high concentrations of HsdR. However, in the absence of structural

information for the subunits of EcoR124, further functional analysis of HsdS-HsdM-DNA interactions remains limited.

1.3.3. HsdM – the modification subunit

HsdM subunit, or simply M, is involved in DNA modification by transferring a methyl group from the donor S-adenosyl-L-methionine (SAM) to the N6 atom of an adenine base within the recognition sequence (68). This process, called methylation, only occurs when HsdM is a part of the M_2S complex, or MTase.

Methylation typically happens when one of the adenines in the recognition sequence is already methylated while the other is not. In other words, the recognition sequence is hemimethylated. Mutations in M.EcoKI can turn it into a *de novo* MTase (69), although the mechanism behind it is not fully understood.

Five crystal structures of HsdM from Type I R-M enzymes have been elucidated (Table 1, Figure 5), mostly from putative R-M systems with a notable exception of M.EcoKI. Only one HsdM structure form *Vibrio vulnificus* has been described in a standalone publication (18), while M.EcoKI and the HsdM from *Bacteroides thetaiotaomicron* were later used for MTase reconstruction based on negative-stain electron microscopy (16).

Structural analysis of the HsdM from *Vibrio vulnificus* (vvHsdM) describes the subunit as consisting of an α-helical N-terminal domain (NTD) and α/β C-terminal domain (CTD) (18). The homologous domains were previously identified for M.EcoKI through partial proteolysis (70). NTD is less conserved across the family members than CTD. According to EM reconstructions of MTase.EcoKI, NTD can act as a dimerization domain between two HsdM subunits, while the extreme C-terminal region of CTD (residues ~470-529) may interact with coil-coiled region of HsdS (16). CTD also interacts with SAM via several residues (Ser289, Leu290, Pro311 and Phe337 in case of vvHsdM, Figure 5) conserved across the family (18). The aromatic rings of Phe199 and Phe312 were found to sandwich the adenine rings, bringing them into a close proximity to the methyl group donor SAM, thus facilitating methyl group transfer (18).

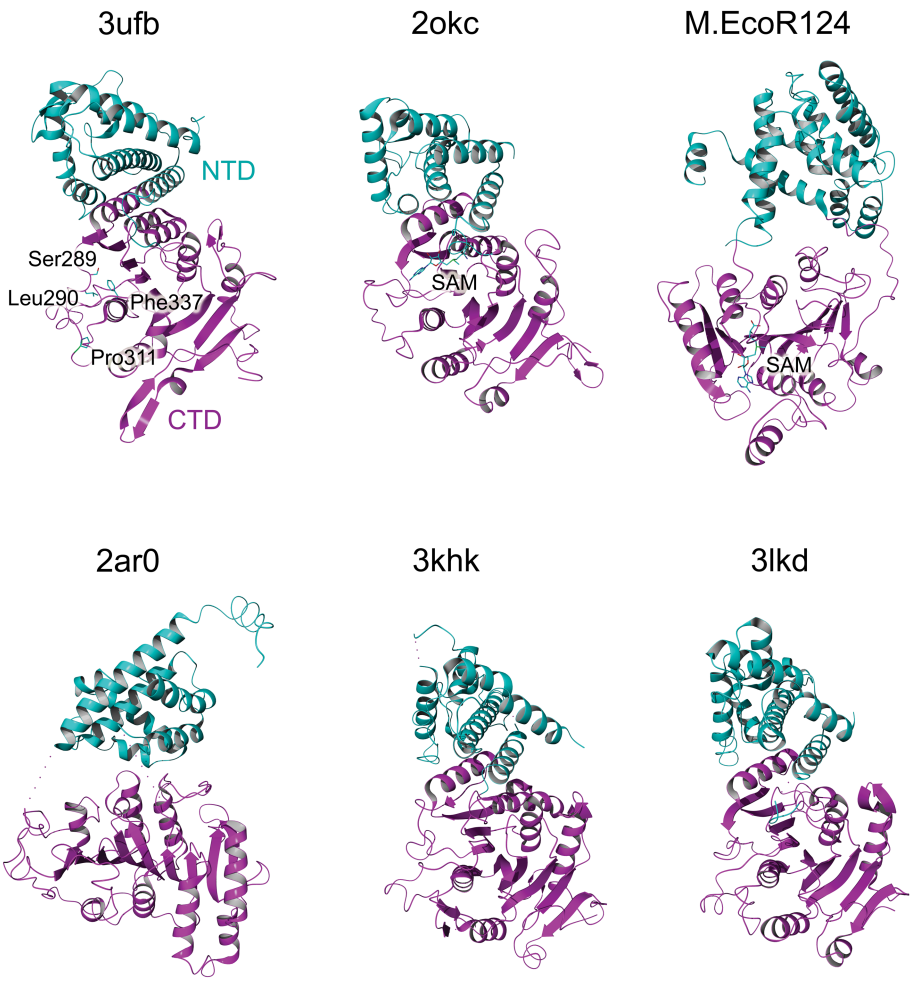

FIGURE 5

3D structures of HsdM subunits. The crystal structures are labeled with their PDB IDs, while M.EcoR124I is a homology model reported by Obarska et el. [ftp://gene-silico.pl/iamb/models/MTases/M.EcoR124I/M.S.EcoR124I.DNA.pdb, (59)]. The N-terminal (NTD) and C-terminal domains (CTD) are in cyan and magenta, respectively. The residues proposed to interact with SAM are marked in the HsdM structure from *Vibrio vulnificus* (3ufb); SAM is shown in the HsdM from *Bacteroides thetaiotaomicron* (2okc) and in the M.EcoR124I homology model. All subunits, except the partial 2ar0, are shown in the same orientation after MUSTANG alignment (61) in YASARA (62).

The HsdM from EcoR124 is a rather large protein of 520 amino acids with molecular weight of ~58 kDa. It is soluble and can be purified as a separate protein, unlike HsdS (58). An early homology model for M.EcoR124 (Figure 5) as a part

of MTase obtained by Obarska *et al.* (59), nevertheless, a crystal structure is required in order to further understand the modification process and HsdM interactions with the other subunits and DNA.

1.3.4. HsdR – the restriction subunit

HsdR subunit is the largest and most intricate constituent of a Type I R-M system. It is responsible for restriction (hence the name, the restriction subunit, or R) and ATP-dependent DNA translocation (hence another name, the motor subunit); both are only exhibited in complex with a DNA-bound MTase.

The first known crystal structures of HsdR came from EcoR124. The WT structure was solved in 2009 (2) followed by four subunits with point mutations (K220R, E220E, K220A and E165H, see Table 1). Around 150 C-terminal residues are not resolved in these structures, for instance the WT structure resolves ~890 out of 1038 residues. The resolved residues fold into four domains: one N-terminal endonuclease, two RecA-like helicase and one helical (Figure 6, top). Crystal structure of a truncated HsdR subunit of a putative Type IB R-M complex from *Vibrio vulnificus* was also reported (15), in which ~600 out of 817 amino acids were resolved, folded into the endonuclease and two RecA-like helicase domains, while the helical domain was missing (Figure 6, bottom).

The endonuclease domain, presenting the typical αβα core, is structurally similar to other nuclease domains, such as those from tRNA endonuclease and the Type II REase EcoRV. In the case of the HsdR from *V. vulnificus*, the sequence identity is very low at below 16% with root-mean-square deviations (rmsd) of slightly below 4 Å when compared to the above-mentioned endonucleases (15). When endonuclease domain of the HsdR from *V. vulnificus* was compared to R.EcoR124, the sequence identity stood at a slightly higher 17%, but with a higher rmsd of 4.6 Å (15). Within the endonuclease domain of R.EcoR124 spanning the residues 13–260, Asp151, Glu165 and Lys167 are proposed to be involved in catalyzing DNA cleavage, while Lys220, closely contacting the ATP moiety, is implicated in coupling translocation and endonuclease activities (2). A similar set of catalytic residues was identified for HsdR from *V. vulnificus* (15), which form a catalytic active site characteristic of PD-(D/E)XK superfamily nucleases (71). Moreover, the motives I, II, III and QxxxY, common to the

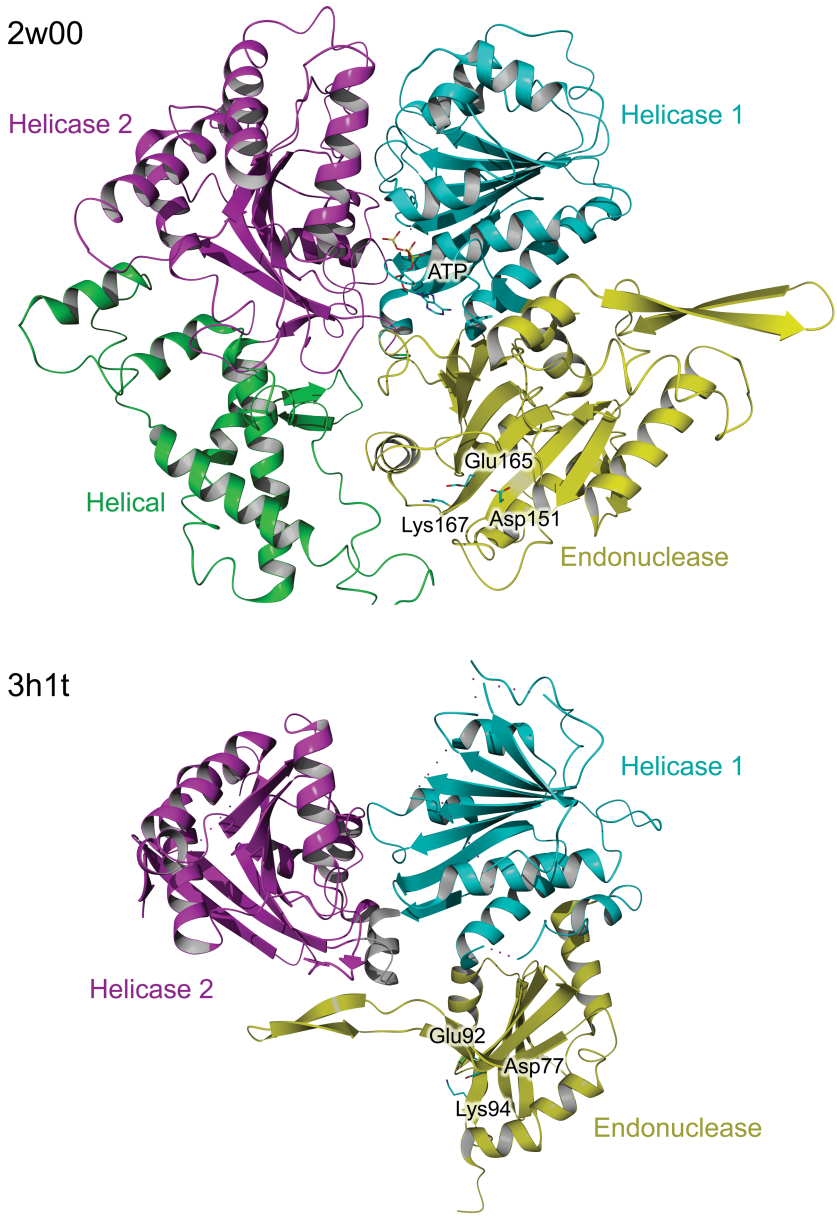

FIGURE 6

Two known crystal structures of HsdR subunits from EcoR124 [top, PDB ID 2W00 (2)] and *Vibrio vulnificus* R-M system [bottom, PDB ID 3H1T, (15)]. The domains are color-coded identically in both structures. The catalytic residues are marked in both structures, while the ATP molecule is labeled in R.EcoR124. The subunits were aligned using MUSTANG (61) in YASARA (62).

superfamily, were also identified within the endonuclease domain. In EcoR124,

mutations in the QxxxY motif decrease DNA binding and the efficiency of DNA cleavage (72), while mutations in catalytic residues not only significantly affect DNA cleavage but also impact translocation (73). Further studies on the QxxxY motif related the alternative interactions of 180s and 220s loops to rotational conformational changes in the endonuclease domain, thus conveying the ATP-ligation status to the complex (74). Extensive mutagenic and computational analysis performed on Lys220 supported the idea that conformational changes in 220s loop are likely to initiate the switch from translocation to endonucleolytic states of the EcoR124 complex (19).

The helicase domains I and II, that encompass the residues 261–461 and 470–731 in case of EcoR124 (2), consist of a parallel β-sheet flanked by an α-helical bundle on either side. Structural alignment revealed an overall fold similar to that of DNA repair helicases, such as RAD25, UvrABC and SNF2/RAD54 (15). Not unlike in the RecA helicases, seven conserved motifs were identified in the helicase domains of HsdR from *V. vulnificus*, four in the helicase domain I and three in the helicase domain II (15). A similar motif composition was characteristic of R.EcoR124 (2). Despite low sequence identity between the domains, they share a similar fold with an rmsd of 3.5 Å over the superimposed Cα atoms for HsdR from *V. vulnificus* (15). Both the helicase domains form a positively charged helicase cleft able to accommodate the size of dsDNA (2). In the crystal structure of R.EcoR124, the ATP moiety contacts the residues Val271, Arg273 and Gln276 in the helicase I and Asp664, Arg688 and Arg691 in the helicase II domains, while the Mg^{2+} ion is coordinated by Lys313, Thr314 and Glu409 (2). In the case of HsdR from *V. vulnificus*, a conserved DECH sequence in the motif II is proposed to bind ATP and Mg^{2+}, analogous to other DExx helicases (15). Despite their names, the helicase domains seem to act as translocases, tracking on one strand of double-stranded DNA without strand unwinding (75).

The helical domain, named after its predominantly α-helical composition, corresponds to residues 732–892 in R.EcoR124 and has no apparent structural relatives or conserved regions (2). Since over ~220 C-terminal residues were cleaved off during the crystallization of HsdR from *V. vulnificus*, the helical domain is absent in the crystal structure (15). It has been a long-standing belief that the C-terminal region is involved in complex formation by binding to MTase (76). Partial proteolysis of the C-terminus of the HsdR form a related

Type I R-M enzyme EcoKI in produced a fragment stable in the presence of ATP but unable to bind to MTase (77). A recent study suggests that the helical domain can mediate DNA translocation, cleavage and ATPase activity via contacts with the helicase 2 domain, thus having a regulatory function (78). A high mobility and flexibility of the C-terminal region seems to be characteristic for HsdR subunits, which is exemplified by the fact that all five R.EcoR124 structures do not resolve ~150 C-terminal residues. The elusive nature of the C-terminal region and its seeming unrelatedness to other proteins made it the main focus of this dissertation.

1.3.5. MTase and pentameric complex

The uniqueness of Type I R-M enzymes lies in their multisubunit nature, when three distinct subunits – HsdS, HsdM and HsdR, described in the previous sections, give rise to several functional complexes. It appears that none of the subunits act separately, only exhibiting their enzymatic potential as part of a larger assembly. Two such complexes are the most important: the methyltransferase (MTase) and the pentameric complex. MTase is formed from one HsdS and two HsdM subunits (M_2S_1 stoichiometry) acting as a site-specific nucleotide methyltransferase. Successive binding of two HsdR subunits to the MTase core forms the pentameric complex ($R_2M_2S_1$ stoichiometry) capable of ATP-dependent DNA translocation and DNA cleavage. In the case of EcoR124, the transient tetrameric complex $R_1M_2S_1$ exists in dynamic equilibrium with $R_2M_2S_1$, with the tetrameric complex being capable of DNA translocation but not cleavage (60).

Only one crystal structure of a Type I R-M complex has been solved to date – the MTase from *Thermoanaerobacter tengcongensis* in DNA-bound form [PDB ID 5YBB, (20)]. Other models of MTase and pentameric complex were obtained by fitting the known crystal structures of individual subunits into 3D density maps generated by negative-strain electron microscopy and other methods. The MTase from EcoKI [PDB IDs 2Y7C and 2Y7H (16)] and DNA bound and free pentameric complexes of EcoR124I [Electron Microscopy Data Bank accession codes 1890 and 1891, (60)] and EcoKI [EMDB accession code 1893, (60)] produced by this method significantly improved the understanding of DNA methylation, translocation and cleavage by Type I R-M enzymes.

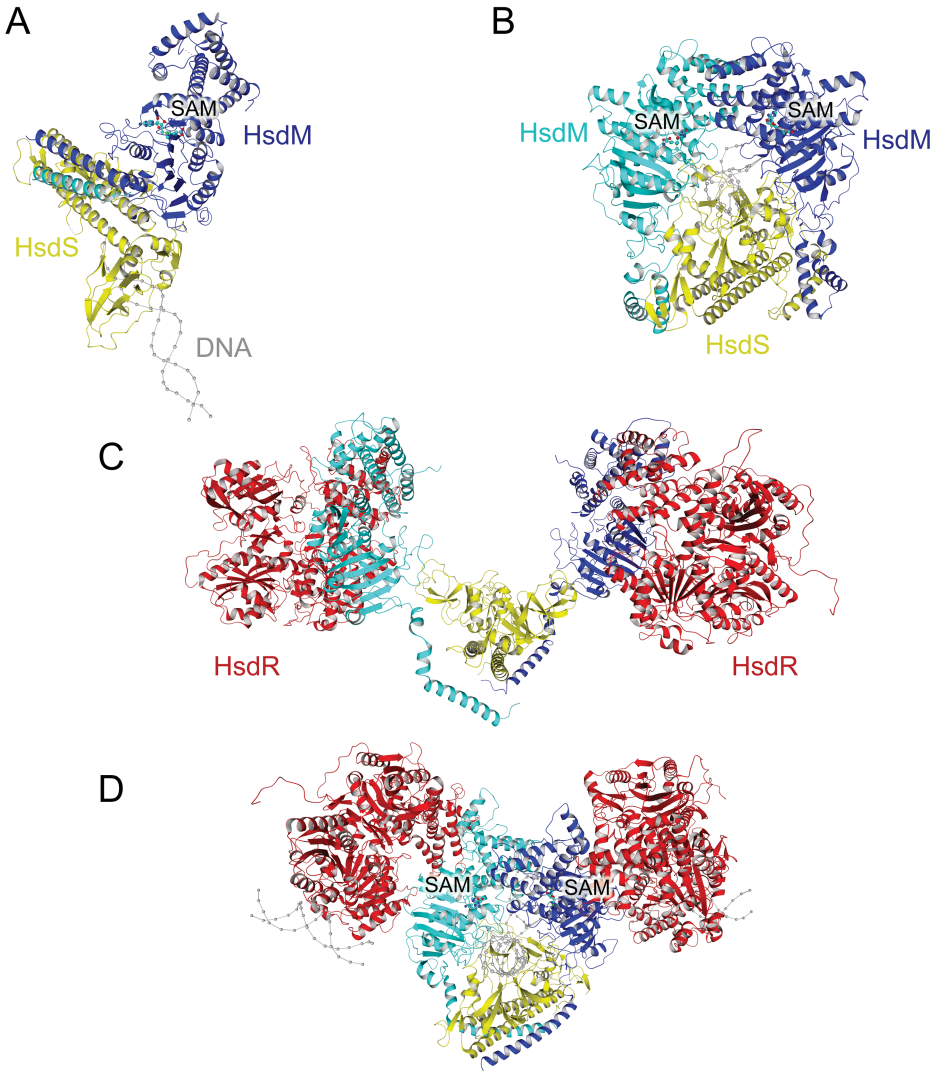

FIGURE 7

3D models of Type I R-M complexes. A. Crystal structure of the open-form MTase from *T. tengcongensis* [PDB ID 5YBB, (20)]. B. Negative stain microscopy reconstruction of the closed-form MTase.EcoKI [PDB IDs 2Y7C, 2Y7H (16)]. Open (C) and closed (D) forms of the EcoR124I pentameric complex; negative stain microscopy reconstruction [ftp://genesilico.pl/iamb/models/RM.TypeI, (55)]. The subunits and DNA have the same colors throughout (HsdS, yellow; HsdM, cyan and blue; HsdR, red; DNA, gray), SAM is labeled where present. Figure prepared in YASARA (62).

A recent crystal structure of the MTase from *T. tengcongensis* (tteMTase) was solved at 3.2 Å in an open conformation. The open form is believed to exist in solution while the closed conformation is triggered by MTase recognizing the

target DNA sequence. Nevertheless, the open form of tteMTase was crystallized in presence of short oligonucleotides, likely representing unspecific DNA affinity of tteMTase (Figure 7A). In contrast, MTase.EcoKI model was obtained in closed conformation (Figure 7B). Two forms of MTase exhibit a significantly different interface between HsdS and HsdM subunits. The subunits of the open-form tteMTase mostly interact via a four-helix bundle formed by two CRs of HsdS (see Figure 1Figure 4 for HsdS structure clarification) and the C-terminal α-helices of both HsdM subunits [Figure 7A, (20)]. To the contrary, each TRD of the HsdS subunit contacts one HsdM subunit in the closed-form model of MTase.EcoKI [B, (16)]. The tteMTase structure, while adding atomic detail to the HsdS-HsdM interface, was suggested to be too crystallographically constrained to represent an actual MTase assembly due to unspecific DNA interactions and the absence of most part of one of the HsdM subunits (79).

According to electron microscopy reconstructions and data from other methods and analogously to the trimeric MTase, the pentameric EcoR124 complex exists in two forms, open and closed, corresponding to DNA-unbound and DNA-bound states (60). 3D models of the open and closed forms are shown in Figure 7, panels C and D, respectively. The closed form is more compact with the MTase core and two HsdR subunits attached on either side, while the open form shows a concerted outward movement of HsdM and HsdR subunits. Similar to MTase, the EcoR124 complex preferably exists in the open form in solution, while the closed form is triggered by HsdS finding the target recognition sequence on DNA. The open form can nonspecifically bind DNA via HsdR subunits (60).

1.3.6. Enzymatic activities – DNA recognition, modification, translocation and restriction

Being a typical Type I R-M complex, EcoR124 exhibits at least four enzymatic activities: [1] recognition of specific target DNA sequence and methylation state readout, [2] transfer of methyl group to the adenine within the recognition sequence, [3] ATP-driven DNA translocation and [4] dsDNA cleavage at a random site. The order of reaction steps depends on the methylation state of the adenines within the recognition sequence (Figure 8). If the adenines in each

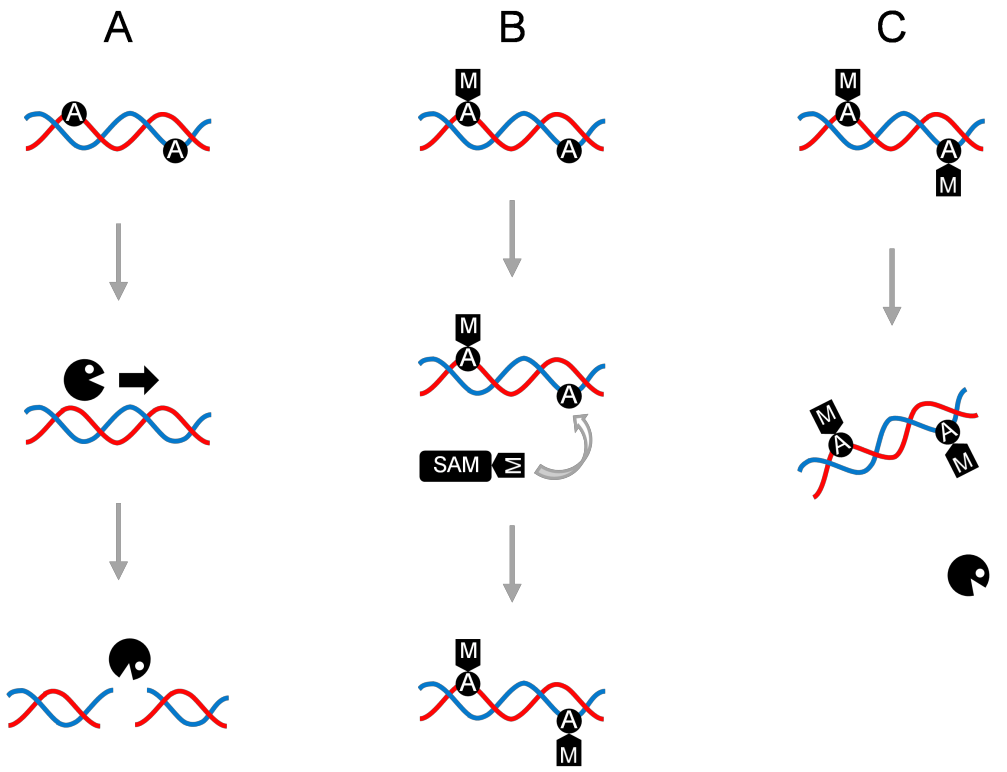

FIGURE 8

Three possible scenarios of EcoR124 enzymatic behavior depending on the methylation state of adenines within the recognition sequence: (A) Both adenines are unmethylated which triggers DNA translocation and cleavage; (B) an adenine in only one DNA strand is methylated, leading to methyl group transfer to the adenine in the daughter strand; (C) adenines in both DNA strands are methylated, causing dissociation of the enzyme from DNA.

DNA strand are unmodified, the DNA molecule is deemed foreign which triggers DNA translocation and subsequent cleavage (step sequence 1→3→4, Figure 8A). If an adenine in only one DNA strand is unmodified (so-called 'hemimethylated state'), the DNA molecule is recognized as self, incurring methyl group transfer from SAM to the N6 atom of the adenine (step sequence 1→2, Figure 8B). The third scenario takes place if both adenines are modified, whereby the complex dissociates from the DNA molecule (step sequence 1→*, Figure 8C). The details of each enzymatic step are given below.

DNA sequence recognition is a key ability of R-M enzymes which allows them to distinguish between self and non-self DNA. The HsdS subunit is responsible

for DNA recognition within the complex. The HsdS of EcoR124I recognizes 5′-GA**A**$(N)_6$R**T**CG-3′ sequence, where R is either A or G, the bases in bold correspond to methylation sites and the complementary 3′-5′ sequence is omitted. Similar to most Type I R-M enzymes, the recognition sequence of EcoR124 is asymmetric, corresponding to 'heterodimeric' nature of the N-terminal and C-terminal TRDs. The length of spacer between two specific parts of the recognition sequence can also vary, giving rise to new specificities. The allelic version of EcoR124I, EcoR124II, recognizes the same DNA sequence except a one-base-pair longer spacer [5′-GA**A**$(N)_7$R**T**CG-3′]. The length of the spacer is connected to the length of CRs (as described in part 1.3.2). Numerous hybrid R-M systems were created by combining TRDs from different HsdS subunits and by varying the length of CRs (65,80,81).

Studies on MTase.EcoR124I showed that MTase has a high binding affinity (K_d ~ 10 nM) for the DNA that contains recognition sequences (82,83), while the ratio of equilibrium binding constants for specific:non-specific binding was ~6000 (83). Methylation interference and competition binding experiments demonstrated a key role of three guanines in the recognition sequence for correct MTase binding (84).

DNA modification is achieved by methyl group transfer from the donor SAM to the adenines within the recognition sequence. The methyl-transfer catalytic residues lie in the CTD of HsdM subunit (see § 1.3.3). EcoR124 produces m6A modification generating N6-methyladenine as a result. EcoR124 mostly acts as a maintenance MTase with a preference for hemimethylated substrates, showing a 100–200-fold increase in reaction rates compared to non-modified DNA (83). Upon DNA binding and prior to methyl transfer, the adenine bases in the recognition sequence are flipped out of DNA backbone (84,85) similar to cytosine base-flipping seen in the crystal structures of a Type II MTase M.HhaI (86).

DNA translocation is a very peculiar feature of Type I R-M enzymes that explains remoteness of their recognition and cleavage sites. DNA translocation by EcoR124 requires ATP (87) and induces formation of DNA loops in both directions since the enzyme pulls DNA through itself (85). ATP hydrolysis, coupled to DNA translocation, is performed by the helicase 1 and 2 domains of the motor subunit HsdR. The helicase domains contain seven conserved motifs typical for

the superfamily 2 (2). Initiation of translocation appear to require a short unwinding of DNA strands observed upon DNA binding (88). Nevertheless, EcoR124 proceeds to translocate dsDNA without unwinding characteristic to many other helicases (89). During DNA translocation by EcoR124, advancement of DNA by one base pair consumes hydrolysis energy of one ATP molecule proceeding in steps of 1–2 bp (90). Both HsdR subunits of EcoR124 can translocate DNA independently; translocation can be (re)initiated and terminated multiple times by HsdR binding and dissociation from the MTase core (91). During translocation, the HsdR subunits of EcoR124 show a measurable turnover only on linear DNA substrates since they probably exit via DNA ends to dissociate from DNA (92). DNA cleavage occurs when translocation is stalled, presumably as a result of changes in DNA topology, collision with an oppositely translocating enzyme or presence of a Holliday junction (93).

DNA restriction, or DNA cleavage, occurs after translocation is stalled and is catalyzed by the endonuclease domain of HsdR. The catalytic triad, Asp151, Glu165 and Lys167, is implicated in endonuclease activity of R.EcoR124 [Figure 6, (2)]. ATP, SAM and Mg^{2+} are required cofactors for DNA cleavage. According to Simons *et al*. (92), HsdR subunits do not show a measurable turnover after cleavage, presumably because they remain tightly bound to the DNA substrate. Alternatively, Bianco *et al.* (94) states that EcoR124I is able to cleave linear DNA substrate continuously added to the reaction, concluding that it is a true enzyme capable of turnover. It is worth noting that both publications heavily rely on cleavage and binding assays *in vitro* where the conditions, including protein and DNA concentration, can be far from physiological, and thus further research is needed to establish the dynamics of cleavage reaction.

Cleavage by EcoR124I occurs without apparent preference for a particular base composition surrounding the restriction site, preferentially generating 5′-overhangs of 3–5 nucleotides in length (95). A single recognition site is sufficient for circular DNA for cleavage to occur, whereas linear substrates require at least two (96). The cleavege site on a linear DNA substrate is located between two recognition sites (96), corroborating the two-enzyme-collision hypothesis.

2. MATERIALS AND METHODS

2.1. Protein production

Production of recombinant proteins is the central step in their structural and biochemical characterization. Typically, this includes vector design, cloning, overexpression and purification stages. Below, a detailed description of protein production is given for MTase, HsdR and its mutant, and pHluorin-HsdR fusion proteins.

2.1.1. MTase

EcoR124I.MTase (Table 3), used in biochemical assays, was expressed from pJS4M vector (82) that contains *HsdS* and *HsdM* genes under control of two independent T7 promoters. *E. coli* strain JM109(DE3) was transformed with plasmid preparation and grown on LB plates containing 100 mg l^{-1} ampicillin. 500 ml 2×YT medium with ampicillin added to 100 mg l^{-1} was inoculated with 5 ml starter overnight culture and cultivated at 37 °C until $OD_{600} \approx 0.4–0.5$ was reached. 1 mM IPTG was then added to induce expression; the culture was grown for another 6 h under vigorous shaking (~200 rpm) and cooled on ice for 30 min. The cells were harvested by centrifugation (4000 g 4 °C for 20 min), washed twice with 40 ml STE buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 8), pelleted and used immediately or stored at -20 °C for up to several months.

Pelleted cells (~1 g) were resuspended in 10 ml lysis buffer [50 mM Tris, 5 mM EDTA, 25%(*w/v*) sucrose, 3 mM DTT, pH 8] and incubated with 1 mg l^{-1} lysozyme for 30 min on ice. All further manipulations were performed at 4 °C. Cell disruption was carried out with a French press, and the lysate was collected by centrifugation (23000 g for 30 mins). Since MTase is a DNA-binding protein, excessive nucleic acids were removed by $(NH_4)_2SO_4$ precipitation. The salt in solid form was being slowly added under constant mixing during ~4 h until saturation was reached. The pellet containing proteins was collected by centrifugation (23000 g for 30 mins), resuspended in 10 ml buffer A (20 mM Tris, 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT, pH 8) and dialyzed against 2 l buffer A overnight.

The dialyzed sample was cleared by centrifugation (23000 g for 2 h), filtered through a sterile syringe Fisherbrand filter (Fisher Scientific, USA), loaded onto

TABLE 3

Biophysical properties of the recombinant proteins and protein complexes obtained in this study.

Protein Name	*Full length, amino acids*	*Molecular weight* (Da)*	*Extinction coefficient* (M^{-1} cm^{-1})*
S.EcoR124I	404	46 178.94	75 415
M.EcoR124I	520	58 013.40	46 550
R.EcoR124I	1038	120 120.28	98 225
MTase (M_2S)	1444	162 169.71	168 515
EcoR124I (R_2M_2S)	3520	402 374.24	365 090
HsdR887del	886	102 664.37	92 265
pHluorin-HsdR705	580	66 798.15	49 405
pHluorin-HsdR867	418	47 778.89	33 475
pHluorin-HsdR887	398	45 442.36	27 975

* calculated in ExPASy (97).

a pre-equilibrated Q Sepharose column (GE Healthcare, USA), washed with buffer A until a stable baseline was reached and eluted with 250 ml NaCl gradient (50–1000 mM). Eluate was collected in 5 ml fractions, which were visualized on a 10%-polyacrylamide SDS gel. The fractions containing HsdS and HsdM subunits were pooled together, concentrated and washed several times with buffer A using a 50 kDa centrifugal filter unit (Millipore, Germany). For the second purification step, the sample was loaded onto a 5 ml HiTrap Heparin affinity column (GE Healthcare, USA), washed with buffer A until the baseline was reached and eluted with 50 ml NaCl gradient (50–1000 mM). 2 ml fractions were collected and checked by SDS-PAGE analysis. The fractions containing HsdS and HsdM subunits were pooled and concentrated on a 50 kDa centrifugal filter unit (Millipore, Germany). Purification procedure was performed on an ÄKTApurifier (GE Healthcare, USA).

Protein concentration was estimated from sample absorbance at 280 and 320 nm wavelength according to Beer-Lambert law:

$$c = \frac{A_{280} - A_{320}}{\varepsilon l},$$

where c is the concentration, A is the absorbance, ε is the molar extinction coefficient and l is the path length; ε was calculated from protein sequence in

ExPASy (97). The absorbance of protein samples was measured on BioPhotometer plus (Eppendorf, Germany).

2.1.2. HsdR and C-terminal deletion mutant

The wild-type and C-terminal deletion mutant HsdR subunits (Table 3) were produced from the pTrcR124 vector commonly used for HsdR overexpression (98). The vector is derived from pTrc99A and contains ampicillin resistance gene and *HsdR* gene expressed from an inducible *lacI* promoter (Table 4). The vector was transformed into *E. coli* strain BL21(DE3) and grown on LB plates with 100 mg l^{-1} ampicillin. 1 l LB medium supplemented with 100 mg l^{-1} ampicillin was inoculated with 10 ml starter culture, grown at 37 °C until $OD_{600} \approx 0.6$ was reached and induced with 1 mM IPTG. After 4–5 h cells were harvested as described above (§ 2.1.1).

Cell paste (~2 g) was resuspended in 20 ml buffer A (20 mM Tris, 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT, pH 8) supplemented with 1 mg l^{-1} lysozyme and incubated for 30 min on ice. The whole purification procedure was performed at 4 °C. Cells were disrupted with a French press, and the lysate was cleared by centrifugation (23000 g for 2 h). The supernatant was filtered through a sterile

TABLE 4

List of oligonucleotides used for construct design and mutagenesis.

Construct	*Oligo name*	*Oligo sequence (5'→3')*
HsdR887del	HsdR-887del-F	GCTGAAGTCTCAG**TAGCCCAATTCGTGTTTTTC**[1]
	HsdR-887del-R	ACGAATTGGGCTA**CTGAGACTTCAGCAAATCG**
	B1027-F	**CCGACATCATAACGGTTCTGGC**
	B629-R	**AGTCCTGCTCGCTTCGCTAC**
pHluorin-HsdR705	HsdR-EcoRI-F	*GAATTC*[2]**TTCCGGGATCTGGAACG**
	HsdR-BamHI-R	*GGATCC*[3]**CTATATTTTTCCGCCTACGCC**
pHluorin-HsdR867	HsdR-867-F	AAAGAATTC**GAGAAATCAACCACTGACTG**
	HsdR-867-R	GGTTGATTTCTC**GAATTCTTTGTATAGTTC**
pHluorin-HsdR887	HsdR-887-F	AAAGAATTC**GAAATAAACCTGGATTATATC**
	HsdR-887-R	CAGGTTTATTTC**GAATTCTTTGTATAGTTC**

[1] Binding sequences are in bold, [2] EcoRI recognition sequence, [3] BamHI recognition sequence.

syringe filter (Fisherbrand, Fisher Scientific, USA) and loaded onto a pre-equilibrated Q Sepharose column (GE Healthcare, USA). After washing with buffer A, the proteins were eluted with 250 ml NaCl gradient (50–1000 mM) into 5 ml fractions and analyzed by SDS-PAGE. Protein was concentrated on 100 kDa centrifugal filter units (Millipore, Germany); the concentration was measured as described above (§ 2.1.1). Purification was performed on ÄKTApurifier (GE Healthcare, USA).

Deletion mutant of HsdR subunit was produced by means of one-step PCR-based method (99) which eliminates the need to perform the standard digestion-ligation protocol. The primers (Table 4) for this method are designed to have 5′-complementary regions circularized *in vivo* by recombination in *E. coli*. A typical 20 µl PCR mix consisted of 1–2 ng pTrcR124 template, 0.5 µM forward and reverse primers, 200 µM dNTP mix and 0.4 U Q5 high-fidelity DNA polymerase in Q5 reaction buffer (New England Biolabs, US). PCRs were performed on Eppendorf Mastercycler gradient (Eppendorf, Germany). Normally, annealing temperatures were 3 °C above the melting temperature of the lower melting temperature primer and the extension time was 30 s per 1 kb of the PCR product. PCR products were treated with 1 µl DpnI (New England Biolabs, USA) to digest the methylated DNA templates (99). If the PCR products were expected to be significantly shorter than the original template, excising the corresponding band from a 1% agarose gel and cleaning it with QIAquick Gel Extraction kit (Qiagen, The Netherlands) had proved to be a viable alternative. After either treatment, 1–5 µl DNA was used to transform XL1-Blue or DH5α *E. coli* competent cells.

Typically, 5–10 transformants were screened by colony PCR with forward and reverse primers (Table 4) upstream and downstream from the expected mutation site. The presence of deletion was verified on a 1% agarose gel using WT pTrcR124 as a control. The colonies with deletion present were grown overnight in 5 ml LB medium supplemented with ampicillin. Next morning, plasmid DNA was isolated with Zyppy Plasmid Miniprep Kit (Zymo Research, USA) and the whole *HsdR* gene sequenced to confirm absence of additional mutations.

2.1.3. pHluorin-HsdR fusion constructs

In order to aid expression, purification and crystallization of HsdR's C-terminal domain, three fusion constructs of varying length with pH-sensitive GFP variant pHluorin (100) were developed. pUC-Kan-pH vector was used for cloning as previously described in *Paper I*. The resulting constructs contained 6×His tag, *pHluorin* gene, N-terminal fusion tag (coding for Glu and Phe) and *HsdR* sequences coding for amino acids 705–1038, 867–1038 and 887–1038.

Expression was carried at 37 °C in 2 l LB medium supplemented with kanamycin. Purification was a two-step procedure including affinity and ion-exchange chromatography. A detailed description is provided in *Paper I*.

2.2. X-ray crystallography

The aim of X-ray crystallography is obtaining a protein structure described as a list of coordinates of the atoms that constitute the protein of interest. Protein structure is modeled by fitting atom coordinates into electron density map. The method is based on the electrons within a crystal causing X-rays to diffract in a predictable manner forming a diffraction pattern. It is therefore necessary to obtain the protein of interest in crystal form. This end is achieved by changing the protein solubility by varying its concentration in presence of different organic and inorganic compounds – the process of crystallization. X-ray diffraction data is then collected using a high-energy X-ray source, and the data is processed using several software packages. This process is described below for pHluorin-HsdR887 fusion protein.

2.2.1. Crystallization

Glyphon crystallization robot (Art Robbins Instruments, USA) was used to perform initial screening in MRC 2-well crystallization plates (Hampton Research, USA) using the sitting-drop vapor-diffusion technique. Several commercial crystallization screens were used depending on availability and the protein – Morpheus, Morpheus II, MIDAS, Structure Screen (Molecular Dimensions, UK), Index Screen, PEG/Ion (Hampton Research, USA), PEGs Suite, PACT Suite (Qiagen, The Netherlands). Typically, the protein in the concentration range 6–20 mg ml^{-1} was mixed with the precipitant in ratios of 1:1 and 2:1 in a

0.4 or 0.6 µl drop volume with reservoir volume 70 µl filled with either the precipitant or 1.5 M NaCl.

After the initial screening, the conditions where crystals formed were repeated manually in a larger 5 µl drop and 700 µl reservoir on 24-well CombiClover crystallization plates (Jena Bioscience, Germany). The conditions were optimized by varying protein concentration and/or that of compounds in the precipitant; supplementary chemical compounds from the Additive Screen (Hampton Research, USA) were used for further optimization where appropriate.

The plates were mostly incubated at 4 °C and occasionally at room temperature. At first, crystal growth was monitored daily and with increasing intervals after the first week. XZX9 stereomicroscope with the attached PEN E-PL3 camera (Olympus, Japan) were used to observe and photograph the crystals.

2.2.2. Data collection and processing

Preliminary testing of crystals was performed on the home source of X-rays – Venture D8 diffractometer equipped with Photon II CPAD detector (Bruker, USA). The home source allowed to test cryoconditions and eliminate the possibility of mistaking crystals of other compounds for protein crystals. The most common cryoprotectors used were PEG 3350, glycerol and ethylene glycol.

Synchrotron radiation was still required to collect X-ray diffraction data. The data were collected at two beamlines: P13 operated by EMBL at the PETRA III X-ray radiation source at the DESY campus (Hamburg, Germany) and BL14.1 operated by the Helmholtz-Zentrum Berlin at the BESSY II electron storage ring (Berlin-Adlershof, Germany). Crystals were mounted in LithoLoops (Molecular Dimensions, England) and flash-cooled in liquid nitrogen.

XDS software (101) was used for data processing, while space group determination and scaling was performed in *POINTLESS* and *SCALA* (102) from *CCP4* package (103). Matthews coefficient (104) was determined using *MATTHEWS_COEF* from *CCP4*.

2.2.3. Structure solution and refinement

Structures were solved by molecular replacement using *MOLREP* version 11.5.05 (105) and *PHASER* version 2.7.17 (106) from *CCP4* package (103). Typically, the structure with the closest sequence identity available in PDB was used as a molecular replacement template. *BUCCANEER* autobuild pipeline version 1.5 (107) from *CCP4* was used to trace missing parts of the structure.

REFMAC version 5.8 (108) from *CCP4* and *WinCoot* version 0.8.7 (109) were used for automated and manual structure refinement, respectively. Manual refinement focused on overall fit and geometry. Residues with no or poor electron density around sidechains were modeled to the Cβ atoms.

2.3. *In vitro* biochemical assays

Purified proteins were subjected to biochemical assays *in vitro* to assess their biological activity and ability to assemble into the pentameric complex. DNA restriction is the primary activity the complex possesses and implies that the complex is fully biologically functional. EMSA allows to see the ability of HsdR subunit to bind to MTase in presence of DNA.

2.3.1. Endonuclease assay

Endonuclease assay was designed to test DNA cutting activity of mutant HsdRs in comparison to wild-type protein. The assay was performed as described previously (98). Typically, 150-250 µl reaction mix consisted of 5 nM circular DNA substrate pCDF30 (67) with one EcoR124I recognition site, 40 nM MTase, 240 nM HsdR, 4 mM ATP and 0.2 mM SAM (*S*-adenosyl methionine) in 1× NEB2 Buffer (New England Biolabs, US).

All components except ATP were mixed at room temperature, equilibrated at 37 °C and ATP was added to initiate the reaction. Aliquots of 20 µl were taken at each time point and an equal volume of stop solution (2% SDS, 0.1 M EDTA, 10% glycerol, 0.1% bromophenol blue) was added followed by a 5 min incubation at 65 °C to stop the reaction. The samples were visualized on 1% TAE agarose gels.

2.3.2. Electrophoretic mobility shift assay

Binding of EcoR124I complex to DNA was assessed *in vitro* following the retardation in the electrophoretic mobility of a radioactively labeled 30-mer duplex oligonucleotide (5′-CGTGCA**GAA**TTCGAG**GTCG**ACGGATCCGGG-3′, EcoR124I recognition site in bold). Equimolar amounts of complimentary oligonucleotides were annealed, labeled with [γ^{32}P]-ATP by T4 polynucleotide kinase (New England Biolabs, US) and purified with QIAquick Nucleotide Removal Kit (Qiagen, The Netherlands). Binding reactions were performed in 10 µl buffer containing 50 mM Tris pH 8, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT and 10% (*v/v*) glycerol. 10 nM DNA substrate was mixed with varying amounts of MTase and wild-type or mutant HsdR and incubated for 10 min at room temperature. DNA and protein complexes were separated on a 6% non-denaturing polyacrylamide gel run in TAE buffer at 4 °C and 100 V. After being vacuum-dried for 30 min at 80 °C on a 583 Gel dryer (Bio-Rad Laboratories, USA), the gel was visualized on a Molecular Imager FX (Bio-Rad Laboratories, USA).

2.4. *In vivo* bacteriophage assay

The restriction activity *in vivo* for the wild-type and mutant HsdRs was measured by analyzing the efficiency of plating (EOP) in a canonical bacteriophage plating assay. *E. coli* JM109(DE3) strain (110) was chosen for its lack of *RecA* and restriction genes. Bacteriophage λ_0 was cultivated and its titer determined as described previously (111). The plasmids pACMS expressing EcoR124II MTase and pKF650 expressing all three subunits of EcoR124II (58) were used for positive and negative complementation assays, respectively. 0.5 ml fresh overnight culture was mixed with 3 ml soft agar medium (LA containing 0.6% agar) and spread on LA plates with corresponding antibiotics. Phage dilutions (10^2 to 10^6 pfu ml^{-1}) were then dropped on the plates followed by overnight incubation at 37 °C. The average EOP was determined for wild-type and mutant HsdRs from at least three independent experiments.

3. RESULTS AND DISCUSSION

3.1. Summary and Outline

The main goal of this study is carrying out a detailed analysis of the C-terminal domain of the HsdR subunit from the Type I restriction-modification enzyme EcoR124 using methods of structural and molecular biology, biochemistry, bioinformatics and molecular dynamics simulations. Results chapter is based on three papers (two published and one manuscript) that are put in chronological order and preceded by experiments that were not included in the publications.

§ 3.2 is mainly based on *Paper I* and deals with construct design (§ 3.2.1), expression, purification, crystallization trials and X-ray data analysis (§ 3.2.2) of pHluorin-HsdR fusion proteins. Fusion proteins were designed to aid expression and crystallization of the C-terminal domain of HsdR, since the C-terminal domain could not be expressed on its own. Three pHluorin-fusion constructs using various parts of HsdR's C-terminus were designed and expressed, although only the shortest construct yielded diffracting crystals. The X-ray diffraction data for this shortest fusion protein, pHluorin-HsdR887, was collected and processed at a resolution of 2.45 Å. The pHluorin-HsdR887 crystals belonged to the orthorhombic space group $C222_1$ with unit-cell parameters a = 83.42, b = 176.58, c = 126.03 Å, $\alpha = \beta = \gamma = 90.00°$ and two molecules in the asymmetric unit (V_M = 2.55 $Å^3$ Da^{-1}, solvent content 50.47%).

Crystal structure of the C-terminal domain, the cornerstone of this study, is described in ***§ 3.3***. The main part of this chapter is based on *Paper II*, that deals with in-depth structural, biochemical and bioinformatical treatment of the C-terminal domain (§ 3.3.5) while preceding sections (§§ 3.3.1–3.3.4) explain the experiments that were not included in the publications. Solving crystal structure of the C-terminal domain (PDB ID 5J3N) and re-refining the existing full-length crystal structure of HsdR using known structural information (PDB ID 6H2J) is also provided in *Paper II*. Based on a re-refined structure and that of the C-terminal domain, the first full-length HsdR model is built and analyzed further. C-terminal domain appears to represent a novel six-helix bundle according to Dali server analysis. Calculations of common molecular surface of C-terminal domain and other four domains indicate that most interactions occur with the helical and helicase 2 domains (142.86 $Å^2$ and 39.67 $Å^2$, respectively). Electrostatic surface potential calculations carried out for the full-length HsdR model

show a positively-charged groove between the endonuclease and C-terminal domains that extends the DNA-binding cleft between the two helicase domains. Results of both *in vivo* and *in vitro* tests with C-terminal deletion mutant HsdR indicate that loss of the C-terminus prevents HsdR from binding to MTase and DNA in accordance with its proposed role.

Using known and putative HsdR sequences, conserved regions within the C-terminal domain were identified in *Paper II*. The domain is poorly conserved among known HsdR subunits (~23-27%) which hinders multiple sequence alignment across the subfamilies, therefore ~18000 putative HsdR sequences were analyzed for presence of identified motifs. Four conserved motifs in the C-terminal domain were tentatively assigned: ^{887}E*X*N*X*DYIL^{894}, ^{925}R*X*K*XX*L*XXX*FI^{935}, ^{996}G—^{1004}P*XX*S^{1007} and ^{1016}KK*XXXXX*K^{1023}. The latter could be a good candidate for interaction with MTase.

Paper II goes to conclude that the motor subunit is comprised of five structural and functional domains and the fifth, C-terminal domain reveals a novel fold characterized by four conserved motifs in subfamily IC of the Type I R-M systems, essential for proper complex assembly and probably involved in DNA binding.

§ 3.4 is based on *Paper III* and presents a molecular dynamics (MD) study of the full HsdR with the C-terminal domain. MD simulations were run at three different temperatures (280, 300 and 340 K) for the full-length HsdR in apo- and holo-forms. DSSP analysis showed that the C-terminal domain is stable and keeps its secondary structure throughout the simulations length. Principal component analysis highlighted the correlated motion of the C-terminal domain with the endonuclease domain thus strengthening its possible role in binding and positioning of DNA close to the catalytic site on the endonuclease domain. The study also suggests the relevance of the C-terminal domain in signal transfer from the ATP binding pocket to the catalytic residues through oscillatory engagement of the endonuclease domain with ATP.

3.2. Solving crystal structure of C-terminal domain

The idea of using a fusion construct to aid expression, purification and crystallization has been implemented with a varying degree of success. Here, we constructed three fusion proteins with the aim of solving crystal structure of the C-terminal domain of HsdR. Other standard approaches had been attempted beforehand, such as improving crystallization outcomes with the whole HsdR structure and expressing C-terminal domain on its own. The structures of

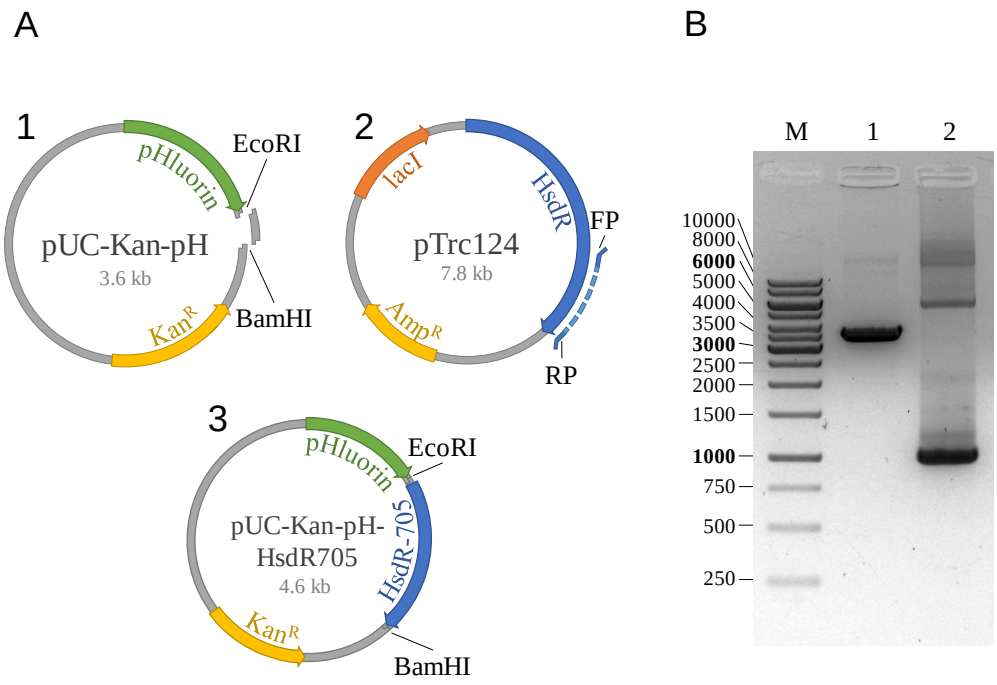

FIGURE 9

Cloning of the C-terminal fragment of *HsdR* coding for amino acids 705–1038 into pUC-Kan-pH vector.

(A) Schematic representation of cloning procedure: [1] the vector digested with EcoRI and BamHI; [2] the C-terminal fragment of HsdR amplified with forward (FP) and reverse (RP) primers flanked by EcoRI and BamHI recognition sequences; [3] "sticky" ends of the vector and insert are ligated.

(B) 1% agarose gel with the digested vector (3.6 kb, lane 1) and PCR products of the C-terminal HsdR fragment (1.0 kb, lane 2). Lane M contains GeneRuler 1 kb DNA ladder (Thermo Scientific, USA) with sizes marked in bp.

HsdR.EcoR124I and four mutants available from PDB (Table 1) all lack atomic-level information in the C-terminal domain, most likely due to its flexibility. Attempts to crystallize wild-type protein yielded crystals with largely similar morphology and diffracting ability (data not shown). The pTrcR124 vector with HsdR's N terminus deleted exhibited low transformation efficiency and no discernable expression in our experiments (data not shown). It was therefore decided to use a GFP variant pHluorin (100) as an expression partner for the C-terminal domain of HsdR. Below, we describe the whole process of obtaining

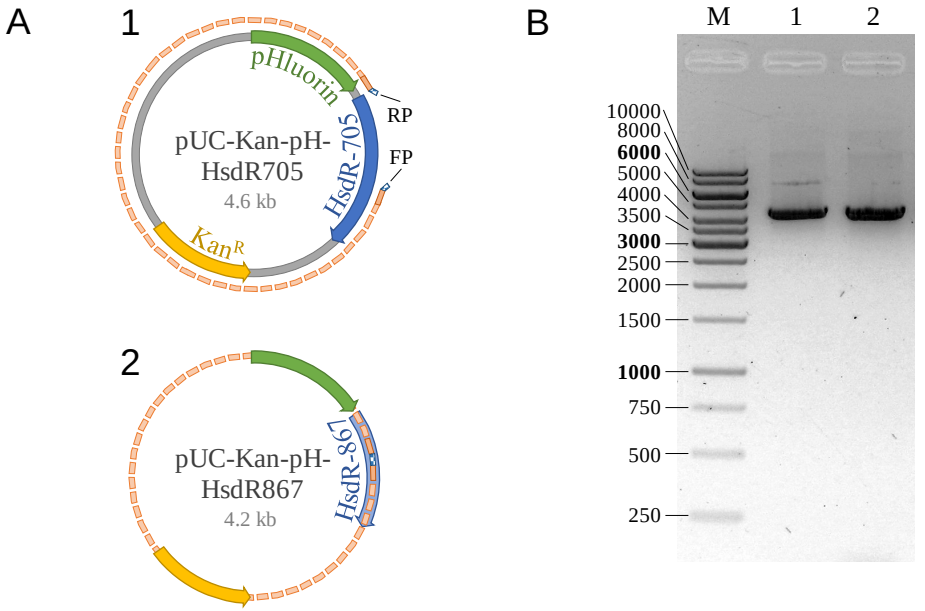

FIGURE 10

Obtaining shorter pHluorin-HsdR867 and pHluorin-HsdR887 fusion constructs from pHluorin-HsdR705.

(A) The scheme of one-step cloning procedure: [1] amplifying the whole vector with HsdR-867-F (FP) and HsdR-867-R (RP) primers that have overlapping regions shown in checkered pattern; [2] linear PCR products transformed into *E. coli* followed by *in vivo* recombination. Analogous cloning was performed for pHluorin-HsdR887.

(B) 1% agarose gel with the whole-plasmid PCR products for pHluorin-HsdR867 (4.2 kb, lane 1) and pHluorin-HsdR887 (4.1 kb, lane 2) construction. Lane M contains GeneRuler 1 kb DNA ladder (Thermo Scientific, USA) with sizes marked in bp.

the crystal structure of the C-terminal domain, beginning with vector design and ending with X-ray data processing and structure solution.

3.2.1. Construct design

pUC-Kan-pH vector (*Paper I*, § 3.2.2) already carrying N-terminal 6×His tag and *pHluorin* gene followed by a multiple cloning site was used to insert *HsdR* sequence conding for amino acids 705–1038. The vector and *HsdR* sequence amplified from pTrcR124 with primers HsdR-EcoRI-F and HsdR-BamHI-R containing EcoRI and BamHI recognition sequences (Table 4) were digested with EcoRI and BamHI restriction enzymes. Expected size for the vector after digestion was ≈3.6 kb and that of PCR products was ≈1.0 kb (Figure 9). The vector and insert were then ligated and transformed into XL1-Blue *E. coli* competent cells. Plasmid DNA was extracted from a few transformants and verified by colony PCR and sequencing as described for HsdR (§ 3.3.2).

Shorter fusion constructs pUC-Kan-pH-HsdR867 and pUC-Kan-pH-HsdR887 were produced from pUC-Kan-pH-HsdR705 by whole-plasmid amplification using corresponding primers (Table 4) as described in methods (§ 2.1.3). The PCR products after whole plasmid amplification were ≈4.2 and 4.1 kb for pUC-Kan-pH-HsdR867 and pUC-Kan-pH-HsdR887 respectively (Figure 10).

3.2.2. pHluorin-assisted expression, purification, crystallization and X-ray diffraction data analysis of C-terminal domain of HsdR

This chapter is based on *Paper I*: Grinkevich P., Iermak I., Luedtke N.A., Mesters J.R., Ettrich R., Ludwig J. (2016) pHluorin-assisted expression, purification, crystallization and X-ray diffraction data analysis of the C-terminal domain of the HsdR subunit of the Escherichia coli type I restriction-modification system EcoR124I. *Acta Cryst. F***72**, 672-676.

ABSTRACT

The HsdR subunit of the type I restriction-modification system EcoR124I is responsible for the translocation as well as the restriction activity of the whole complex consisting of HsdR, HsdM and HsdS subunits, and while crystal structures are available for wild type and several mutants, the C-terminal domain comprising approximately 150 residues was not resolved in any of those structures. Here, three fusion constructs with the GFP variant pHluorin developed to overexpress, purify and crystallize the C-terminal domain of HsdR are reported. The shortest of the three encompassed HsdR residues 887–1038, and yielded crystals that belonged to the orthorhombic space group $C222_1$ with unit-cell parameters a = 83.42, b = 176.58, c = 126.03 Å, α = β = γ = 90.00° and two molecules in the asymmetric unit (V_M = 2.55 $Å^3$ Da^{-1}, solvent content 50.47%). X-ray diffraction data were collected to a resolution of 2.45 Å.

3.3. Structure and function of C-terminal domain

3.3.1. MTase expression and purification

EcoR124I MTase is a tetrameric complex that consists of one HsdS and two HsdM subunits. While HsdM can be easily purified as a standalone protein, HsdS has a very low solubility (58). Both subunits therefore need to be concurrently expressed. Both one and two-plasmid expression systems were employed at different times to overexpress the subunits as described in methods.

Since MTase tends to stay DNA bound and HsdS has low solubility, the non-native buffer conditions were applied during cell lysis followed by DNA removal by $(NH_4)_2SO_4$ precipitation. Ion-exchange chromatography was then employed to purify out most of the cellular proteins. Affinity chromatography on a heparin column was used as a final step. The aim here was to let the subunits bind and assemble on the column and get rid of the misfolded portion, so that the final product mostly contained the physiologically relevant assembled trimeric MTase. Ion exchange and heparin affinity purification steps are shown in Figure 11. Typically, MTase was concentrated to 10–80 μM concentration,

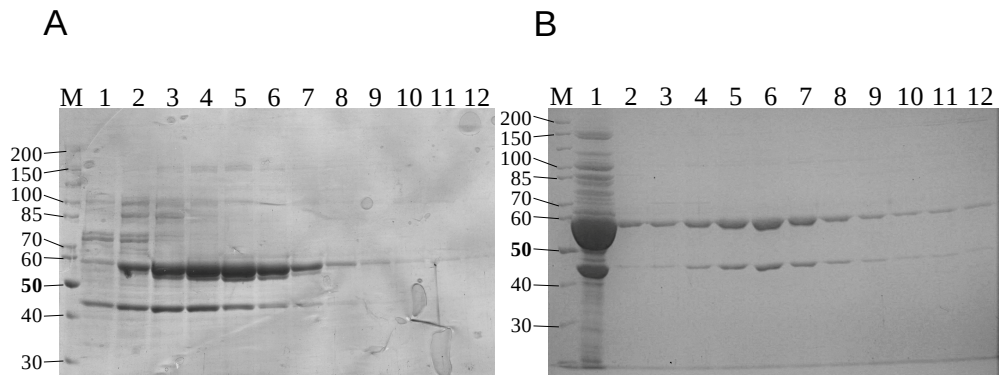

FIGURE 11

Two-step purification of MTase. First, ion-exchange chromatography was used (A) whereupon fractions from lanes 3-6 were concentrated. Heparin affinity column was used as a second step (B) and fractions from lanes 4-8 were pooled for further experiments; lane 1 contains concentrated protein from the first purification step. HsdS and HsdM subunits are visible on both gels as think bands of ~46 and 58 kDa. The marker in the M lane is PageRuler Unstained Protein Ladder (Thermo Scientific, USA); the numbers next to the lane represent molecular weight in kDa.

combined with 50%(*w/v*) glycerol and stored at -20 °C. MTase appeared to be very stable in such cryoprotected form.

MTase was typically used to assemble the EcoR124I pentameric complex in various biochemical assays, such as DNA endonuclease assay and EMSA, because most biochemical properties of HsdR can only be evaluated in the assembled state.

3.3.2. C-terminal deletion mutant production

The C-terminus of HsdR is required for binding to MTase as it was shown for a related Type I R-M system EcoKI in the limited proteolysis experiments (77). Since a similar role has also been suggested for EcoR124 (68,76), the C-terminal deletion mutant needed to be constructed to test this hypothesis. HsdR residues 887–1038 were removed by means of three consequent PCRs, digested with PtsI and BamHI and ligated back into the pTrcR124 vector.

A schematic representation of cloning procedure is given in Figure 12A. Briefly, two overlapping gene fragments were obtained in two independent PCRs, cleaned and combined in an overlap PCR adding only outer primers. The overlap PCR products were then excised from an 1% agarose gel and treated with PtsI and BamHI. The vector was digested with the same enzymes. The corresponding fragments were excised from the gel, ligated and transformed into *E. coli* competent cells. The agarose gel with the intermediate cloning steps is show in Figure 12B. The plates with transformants were grown and the identity of clones was verified by colony PCR and DNA sequencing as described in methods.

3.3.3. HsdR expression and purification

As a soluble protein native to *E. coli*, HsdR subunit was typically easy to overexpress from pTrcR124 vector. High levels of expression are fostered by the presence of the synthetic *trc* promoter in this vector, while vectors's own copy of *lacI* increases the strong expression response to IPTG.

Both WT and mutant HsdRs were expressed in BL21(DE3) strain of *E. coli* and purified with affinity chromatography. Rather high level of expression resulted in a sufficiently pure protein even after one affinity step (data no shown). HsdR

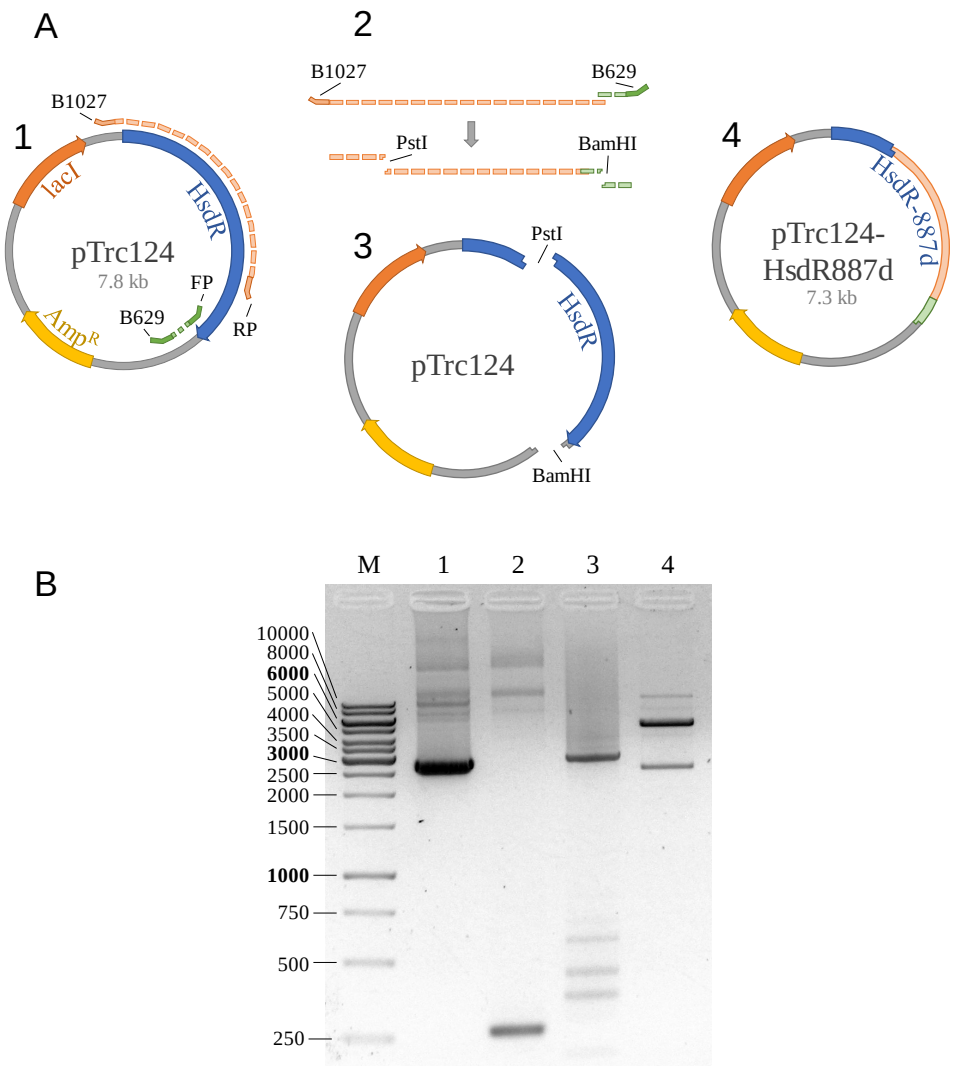

FIGURE 12

Cloning procedure for HsdR 887–1038 deletion mutant (HsdR887del).

(A) Scheme of cloning procedure: [1] - two independent PCRs with HsdR-887del-F (FP) and HsdR-887del-R (RP) primers and a corresponding outer primer; [2] – overlap PCR with the outer primers and subsequent digest of the PCR product and vector [3] with PstI and BamHI; [4] vector with HsdR887del insert after ligation.

(B) 1% agarose gel with two initial PCR products (2.8 and 0.3 kb, lanes 1 and 2), overlap PCR product (3.1 kb, lane 3) and vector digest (5.2 and 2.6 kb) Lane M contains GeneRuler 1 kb DNA ladder (Thermo Scientific, USA) with sizes marked in bp.

was concentrated to a 100–250 μM concentration and either used immediately or stored at -20 °C as a 50%(*w/v*) glycerol stock for up to a few months. Better results in biochemical assays were obtained with a freshly purified protein.

3.3.4. Endonuclease assay with C-terminal deletion mutant

In order to test C-terminal domain's involvement in functioning of the whole complex, the C-terminal domain deletion HsdR mutant was prepared as described above. This mutant HsdR was subjected to *in vivo* and *in vitro* assays as described in *Paper II* (§ 3.2.2). *In vitro* endonuclease assay was omitted from the paper for brevity and its results are given here. This assay allows to test restriction activity using a DNA substrate with a single EcoR124I recognition site, in this case the circular pCDF30 plasmid. After mixing MTase, HsdR and DNA substrate, aliquots were taken upon addition of ATP at various time points and visualized on TAE agarose gels. Supercoiled, linear and 'nicked' forms of DNA substrate migrate through the gel at different speeds, allowing to distinguish between them. Because no significant amounts of linear DNA are present (Figure 13B), C-terminal deletion mutant does not exhibit endonuclease activity compared to the wild type (Figure 13A). This outcome is corroborated by the results reported in *Paper II*: since the mutant is not capable to assemble with MTase, the lack of endonuclease activity is expected.

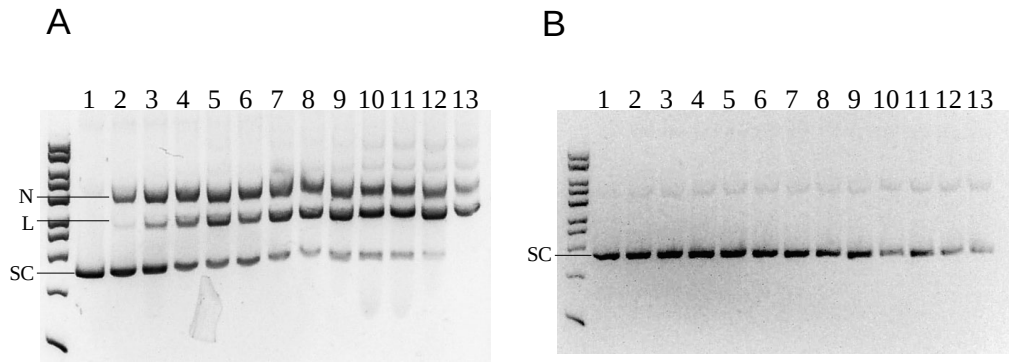

FIGURE 13

In vitro endonuclease assay with wild-type HsdR (A) and C-terminal deletion mutant HsdR (B). Both gels contain aliquots taken during an identical time course experiment where lane 1 is an aliquot before addition of ATP and lanes 2–13 correspond to time points 15, 30, 45, 60, 90, 120, 180, 300, 600, 900, 1800 and 3600 s.

3.3.5. Crystal structure of a novel domain of the motor subunit of the type I restriction enzyme EcoR124 involved in complex assembly and DNA binding

This chapter is based on *Paper II*: Grinkevich P., Sinha D., Iermak I., Guzanova A., Weiserova M., Ludwig J., Mesters J.R., Ettrich, R.H. (2018) Crystal structure of a novel domain of the motor subunit of the Type I restriction enzyme EcoR124 involved in complex assembly and DNA binding. *J Biol Chem*, doi: 10.1074/jbc.RA118.003978.

ABSTRACT

Although EcoR124 is one of the better-studied type I restriction-modification enzymes, it still presents many challenges to detailed analyses because of its structural and functional complexity and missing structural information. In all available structures of its motor subunit HsdR, responsible for DNA translocation and cleavage, a large part of the HsdR C terminus remains unresolved. The crystal structure of the C terminus of HsdR, obtained with a crystallization chaperone in the form of pHluorin fusion and refined to 2.45 Å, revealed that this part of the protein forms an independent domain with its own hydrophobic core and displays a unique α-helical fold. The full-length HsdR model, based on the WT structure and the C-terminal domain determined here, disclosed a proposed DNA-binding groove lined by positively charged residues. *In vivo* and *in vitro* assays with a C-terminal deletion mutant of HsdR supported the idea that this domain is involved in complex assembly and DNA binding. Conserved residues identified through sequence analysis of the C-terminal domain may play a key role in protein–protein and protein–DNA interactions. We conclude that the motor subunit of EcoR124 comprises five structural and functional domains, with the fifth, the C-terminal domain, revealing a unique fold characterized by four conserved motifs in the IC subfamily of type I restriction-modification systems. In summary, the structural and biochemical results reported here support a model in which the C-terminal domain of the motor subunit HsdR of the endonuclease EcoR124 is involved in complex assembly and DNA binding.

3.4. Functional characterization of the fifth domain in HsdR subunit of Type I RM system EcoRI24I: A molecular dynamics study

This chapter is based on *Paper III*: Sinha D., Grinkevich P., Ettrich, R.H. (2018) Functional characterization of the fifth domain in HsdR subunit of Type I RM system EcoRI24I: A molecular dynamics study. *Manuscript*

ABSTRACT

EcoR124I is a Type I restriction-modification system, a multifunctional, multisubunit hetero-pentameric complex with DNA cleavage and ATP dependent DNA translocation activities located on the HsdR subunit. The enzyme recognizes specific DNA target sequence and methylates the hemimetylated DNA but whenever encounters with unmethylated DNA, the complex translocates thousands of bp and cleave the DNA non-specifically at distant sites. In the first crystal structure of HsdR subunit the four domains form a square planar arrangement. The extended part of C terminal surprisingly fitted perpendicular to the plane and folded in the form of fifth domain. Computational modeling including molecular dynamics in combination with crystallography study suggested the possible role of C-terminal domain in binding and positioning of DNA to the catalytic residue during the cleavage activity.

4. CONCLUSIONS

EcoR124 is a plasmid-borne representative of the Type I R-M enzymes that allow prokaryotic cells to discriminate between self and non-self genetic material and degrade the latter, thereby keeping their own integrity. Several enzymatic activities are assigned to three subunits constituting the EcoR124 complex: HsdS recognizes target DNA sequences, HsdM methylates DNA bases and HsdR translocates and cleaves DNA at unpredictable sites. Since dsDNA breaks could be lethal for the host bacterium, triggering genetically programmed cell death akin to apoptosis (112), it is crucial that enzymatic functions of EcoR124 are neatly orchestrated by inter- and intrasubunit communication.

HsdR subunit takes center stage after an unmethylated target DNA sequence has been recognized. This large multidomain protein binds to MTase and initiates translocation and subsequent dsDNA cleavage at a distant site. The X-ray crystal structure of HsdR obtained in 2009 (2) brought a renewed interest in studying HsdR's function. Four domains were identified in this and later HsdR structures – the endonuclease, helicase 1 and 2 and helical. Specific roles were subsequently assigned to each domain. The endonuclease domain is involved in DNA cleavage, the helicase 1 and 2 in ATP-driven DNA translocation and the helical domain has a regulatory function. The C-terminal 150 residues, missing from reported HsdR structures, have remained elusive for structural and functional characterization. This thesis describes the process of obtaining the crystal structure of the C-terminus that turned out to form a separate fifth domain of HsdR and deals with its functional description.

Domains are often connected to each other with flexible regions that permit a range of movement necessitated by their function. In the case of HsdR, the C-terminal domain contributes to protein-protein and protein-DNA interactions and therefore may require substantial movement. Such regions of high mobility are often poorly resolved or absent from crystal structures. Various approaches are implemented to counteract this. Here, we used crystallization chaperone approach to crystallize the C-terminal domain after the initial attempts to overexpress it separately proved unsuccessful. The fusion of C-terminal regions of various length with a ratiometric GFP variant pHluorin led to successful overproduction, purification and crystallization of one of the fusions containing HsdR residues 887–1038. X-ray data were collected and processed for optimized crystals that diffracted to a 2.45 Å resolution. Using GFP as a template for molecular

replacement provided phase information sufficient to obtain an electron density map for the HsdR part and retrace its amino acids.

The resulting crystal structure of the C-terminal domain was used as a search model with the wild-type HsdR data allowing to position the domain with respect to the rest of the structure. Using computational modeling techniques to fill in missing regions and combining side-chain information from both the pHluorin fusion and wild-type structures, we produced the most complete model of HsdR subunit to date. Structurally, the C-terminal domain is a novel six-helix fold comprised of two three-helix bundles stacked at an angle and connected with a short linker conserved in the subfamily IC. A number of the C-terminal domain's interdomain contacts were observed with the helical and helicase 2 domains.

EMSA was used to confirm the role of C-terminus in DNA binding. For this purpose, the C-terminal deletion mutant of HsdR was produced. It showed no affinity to MTase-DNA complex. Additionally, no restriction activity was observed *in vivo* for the mutant. It also did not have a significant negative impact on the wild type binding ability in negative complementation, confirming that the lack of the C-terminal domain has a severe negative impact on HsdR binding the MTase-DNA complex. The mutant was also not able to cleave a circular DNA substrate *in vitro*.

These experiments led us to suppose that the C-terminal domain's involvement in the HsdR–MTase–DNA binding interface. We investigated it further with bioinformatic analysis. Owing to a low amino acid conservation in the C-terminal domain between the Type I subfamilies, we mined the putative HsdR subunits to produce a relevant multiple alignment. Four motifs with a higher conservation level were tentatively assigned in the C-terminal domain, with ^{1016}KK*XXXXX*K^{1023} being a good candidate for participating in binding to MTase. Electrostatic surface potential calculations show a prominent groove lined with positively-charged residues between the endonuclease and C-terminal domains that extends the previously suggested DNA-binding cleft between the two helicase domains. We therefore suppose that the C-terminal domain can help to bring DNA into the correct position through electrostatic interactions.

MD simulation of the full-length HsdR model showed that the protein undergoes major conformational shifts with the C-terminal domain moving as a whole with respect to the other domains. We suggest that this motion can help in correct positioning of DNA near the catalytic residues of the endonuclease domain, thus facilitating DNA cleavage.

In summary, we elucidated certain aspects of the C-terminal domain's involvement in complex assembly and DNA binding by providing the full-length HsdR model through obtaining the first crystal structure of the C-terminal domain and combining information from known crystal structures. Techniques of molecular biology, biochemistry, bioinformatics and computational biology allowed us to characterize the new domain and hopefully bring us one step closer to understanding the functionality of Type I R-M enzymes. Further work needs to be done to decode the involvement of particular residues in the C-terminal domain in assembly of the EcoR124 complex and in DNA interactions.

5. REFERENCES

1. Youell, J., Firman, K. (2008) EcoR124I: from Plasmid-Encoded Restriction-Modification System to Nanodevice. *Microbiol. Mol. Biol. Rev.* 72(2), 365–377.
2. Lapkouski, M., Panjikar, S., Janscak, P., Kuta Smatanova, I., Carey, J., Ettrich, R., Csefalvay, E. (2009) Structure of the motor subunit of type I restriction-modification complex EcoR124I. *Nature Struct. Mol. Biol.* 16(1), 94–95.
3. Luria, S. E., Human, M. L. (1952) A nonhereditary, host-induced variation of bacterial viruses. *J. Bacteriol.* 64, 557–569.
4. Anderson, E. S., Felix, A. (1952) Variation in Vi-phage II of Salmonell typhi. *Nature* 170, 492–494.
5. Bertani, G., Weigle, J. J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.* 65, 113–121.
6. Arber, W., Dussoix, D. (1962) Host specificity of DNA produced by Echerichia coli. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.* 5, 18–36.
7. Arber, W., Dussoix, D. (1962) Host specificity of DNA produced by Escherichia coli. II. Control over acceptance of DNA from infecting phage lambda. *J. Mol. Biol.* 5, 37–49.
8. Arber, W., Linn, S. (1969) DNA modification and restriction. *Annu Rev Biochem.* 38, 467–500.
9. Meselson, M., Yuan, R. (1968) DNA Restriction Enzyme from E. coli. *Nature* 217, 1110–1114.
10. Linn, S., Arber, W. (1968) Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. *Proc Natl Acad Sci U S A.* 59(4), 1300–1306.
11. Meynell, E., Datta., N. (1966) The relation of resistance transfer factors to the F-factor (sex-factor) of Escherichia coli K12. *Genet. Res. (Cambridge)* 7, 134–140.
12. Hughes, S. G. (1977) *Studies of plasmid encoded restriction and modification systems.* PhD thesis, University of Edinburgh, Edinburgh, United Kingdom.
13. Kim, J. S., DeGiovanni, A., Jancarik, J., Adams, P. D., Yokota, H., Kim, R., Kim, S. H. (2005) 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(9), 3248–3253.
14. Calisto, B. M., Pich, O. Q., Piñol, J. F., Querol, E., Carpena, X. (2005) Crystal structure of a putative type I restriction-modification S subunit from Mycoplasma genitalium. *J Mol Biol.* 351(4), 749–762.
15. Uyen, N. T., Park, S. Y., Choi, J. W., Lee, H. J., Nishi, K., Kim, J. S. (2009) The fragment structure of a putative HsdR subunit of a type I restriction enzyme from Vibrio vulnificus YJ016: implications for DNA restriction and translocation activity. *Nucleic Acids Res.* 37(20), 6960–6969.
16. Kennaway, C. K., Obarska-Kosinska, A., White, J. H., Tuszynska, I., Cooper, L. P., Bujnicki, J. M., Trinick, J., Dryden, D. T. (2009) The structure of M.EcoKI Type I DNA

methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res.* 37(3), 762–70.

17. Gao, P., Tang, Q., An, X., Yan, X., Liang, D. (2011) Structure of HsdS subunit from Thermoanaerobacter tengcongensis sheds lights on mechanism of dynamic opening and closing of type I methyltransferase. *Plos One* 6, e17346.
18. Park, S. Y., Lee, H. J., Song, J. M., Sun, J., Hwang, H. J., Nishi, K., Kim, J. S. (2012) Structural characterization of a modification subunit of a putative type I restriction enzyme from Vibrio vulnificus YJ016. *Acta Cryst. D* 68, 1570–1577.
19. Csefalvay, E., Lapkouski, M., Guzanova, A., Csefalvay, L., Baikova, T., Shevelev, I., Bialevich, V., Shamayeva, K., Janscak, P., Kuta Smatanova, I., Panjikar, S., Carey, J., Weiserova, M., Ettrich, R. (2015) Functional Coupling of Duplex Translocation to DNA Cleavage in a Type I Restriction Enzyme. *Plos One* 10, 28700e.
20. Liu, Y. P., Tang, Q., Zhang, J. Z., Tian, L. F., Gao, P., Yan, X. X. (2017) Structural basis underlying complex assembly and conformational transition of the type I R-M system. *Proc Natl Acad Sci U S A* 114(42), 11151–11156.
21. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., Phillips, D. C. (1958) A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* 181(4610), 662–666.
22. Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S. K., Dryden, D. T., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31(7), 1805–1812.
23. Roberts, R. J., Vincze, T., Posfai, J., Macelis, D. (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 43(Database issue), D298–D299.
24. Horiuchi, K., Zinder, N. D. (1972) Cleavage of bacteriophage fl DNA by the restriction enzyme of Escherichia coli B. *Proc Natl Acad Sci U S A.* 69(11), 3220–3224.
25. Loenen, W. AM., Dryden, D. TF., Raleigh, E. A., Wilson, G. G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Res.* 42(1), 20–44.
26. Pingoud, A., Wilson, G. G., Wende, W. (2014) Type II restriction endonucleases – a historical perspective and more. *Nucleic Acids Res.* 42(12), 7489–7527.
27. Hadi, S. M., Bächi, B., Iida, S., Bickle, T. A. (1983) DNA restriction-modification enzymes of phage P1 and plasmid p15B. Subunit functions and structural homologies. *J. Mol. Biol.* 165, 19–34.
28. Yuan, R., Hamilton, D. L., Hadi, S. M., Bickle, T. A. (1980) Role of ATP in the cleavage mechanism of the EcoP15 restriction endonuclease. *J. Mol. Biol.* 144, 501–419.
29. Reiser, J., Yuan, R. (1977) Purification and properties of the P15 specific restriction endonuclease from Escherichia coli. *J. Biol. Chem.* 252, 451–456.

30. Meisel, A., Bickle, T. A., Krüger, D. H., Schroeder, C. (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* 355, 467–469.

31. Meisel, A., Mackeldanz, P., Bickle, T. A., Krüger, D. H., Schroeder, C. (1995) Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO J.*(14), 2958–2966.

32. Rao, N. D., Dryden, D. TF., Bheemanaik, S. (2014) Type III restriction-modification enzymes: a historical perspective. *Nucl. Acids Res* 42(1), 45–55.

33. Loenen, W. AM., Raleigh, E. A. (2014) The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res.* 42(1), 56–69.

34. Bourniquel, A. A., Bickle, T. A. (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie.* 84(11), 1047–1059.

35. Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P. J., Kinney, S. R., Yamada-Mabuchi, M., Xu, S. Y., Davis, T., Pradhan, S., Roberts, R. J., Zheng, Y. (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc Natl Acad Sci U S A.* 108(27), 11040–11045.

36. Webb, J. L., King, G., Ternent, D., Titheradge, A. J., Murray, N. E. (1996) Restriction by EcoKI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J.* 15(8), 2003–2009.

37. Makovets, S., Powell, L. M., Titheradge, A. J., Blakely, G. W., Murray, N. E. (2004) Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease? *Mol Microbiol.* 51(1), 135–147.

38. Korona, R., Korona, B., Levin, B. R. (1993) Sensitivity of naturally occurring coliphages to type I and type II restriction and modification. *J Gen Microbiol.* 139(6), 1283–1290.

39. Korona, R., Levin, B. R. (1993) Phage-mediated selection and the evolution and maintenance of restriction-modification. *Evolution.* 47(2), 556–575.

40. Treangen, T. J., Rocha, E. P. (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7(1), e1001284.

41. Oliveira, P. H., Touchon, M., Rocha, E. PC. (2014) The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* 42(16), 10618–10631.

42. Budroni, S., Siena, E., Dunning Hotopp, J. C., Seib, K. L., Serruto, D., Nofroni, C., Comanducci, M., Riley, D. R., Daugherty, S. C., Angiuoli, S. V., Covacci, A., Pizza, M., Rappuoli, R., Moxon, E. R., Tettelin, H., Medini, D. (2011) Neisseria meningitidis is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc Natl Acad Sci U S A.* 108(11), 4494–4499.

43. Erwin, A. L., Sandstedt, S. A., Bonthuis, P. J., Geelhood, J. L., Nelson, K. L., Unrath, W. CT., Diggle, M. A., Theodore, M. J., Pleatman, C. R., Mothershed, E. A., Sacchi, C. T., Mayer, L. W., Gilsdorf, J. R., Smith, A. L. (2008) Analysis of Genetic Relatedness of

Haemophilus influenzae Isolates by Multilocus Sequence Typing. *J Bacteriol.* 190(4), 1473–1483.

44. Croucher, N. J., Coupland, P. G., Stevenson, A. E., Callendrello, A., Bentley, S. D., Hanage, W. P. (2014) Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nat Commun.* 5, 5471.
45. Furuta, Y., Abe, K., Kobayashi, I. (2010) Genome comparison and context analysis reveals putative mobile forms of restriction–modification systems and related rearrangements. *Nucleic Acids Res.* 38(7), 2428–2443.
46. Dybvig, K., Sitaraman, R., French, C. T. (1998) A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements., 13923–13928.
47. Manso, A. S., Chai, M. H., Atack, J. M., Furi, L., De Ste Croix, M., Haigh, R., Trappetti, C., Ogunniyi, A. D., Shewell, L. K., Boitano, M., Clark, T. A., Korlach, J., Blades, M., Mirkes, E., Gorban, A. N., Paton, J. C., Jennings, M. P., Oggioni, M. R. (2014) A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat Commun.* 5, 5055.
48. Furuta, Y., Namba-Fukuyo, H., Shibata, T. F., Nishiyama, T., Shigenobu, S., Suzuki, Y., Sugano, S., Hasebe, M., Kobayashi, I. (2014) Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS Genet.* 10(4), e1004272.
49. De Ste Croix, M., Vacca, I., Kwun, M. J., Ralph, J. D., Bentley, S. D., Haigh, R., Croucher, N. J., Oggioni, M. R. (2017) Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. *FEMS Microbiol Rev.* 41(Supp_1), S3–S15.
50. Sadykov, M., Asami, Y., Niki, H., Handa, N., Itaya, M., Tanokura, M., Kobayashi, I. (2003) Multiplication of a restriction–modification gene complex. *Mol Microbiol.* 48(2), 417–427.
51. Handa, N., Kobayashi, I. (1999) Post-segregational killing by restriction modification gene complexes: observations of individual cell deaths. *Biochimie.* 81(8-9), 931–938.
52. Weyler, L., Engelbrecht, M., Mata Forsberg, M., Brehwens, K., Vare, D., Vielfort, K., Wojcik, A., Aro, H. (2014) Restriction endonucleases from invasive Neisseria gonorrhoeae cause double-strand breaks and distort mitosis in epithelial cells during infection. *PLoS One.* 9(12), e114208.
53. Bannister, D., Glover, S. W. (1968) Restriction and modification of bacteriophages by R+ strains of Escherichia coli K-12. *Biochem. Biophys. Res. Commun.* 30(6), 735–738.
54. Campbell, I. G., Mee, B. J. (1985) A restriction map of IncFIV plasmid R124. *Plasmid* 14(3), 261–263.
55. Glover, S. W., Firman, K., Watson, G., Price, C., Donaldson, S. (1983) The alternate expression of two restriction and modification systems. *Mol Gen Genet.* 190(1), 65–69.

56. Price, C., Shepherd, J. C., Bickle, T. A. (1987) 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.

57. Price, C., Lingner, J., Bickle, T. A., Firman, K., Glover, S. W. (1989) Basis for changes in DNA recognition by the EcoR124 and EcoR124/3 type I DNA restriction and modification enzymes. *J Mol Biol.* 205(1), 115–125.

58. Patel, J., Taylor, I., Dutta, C. F., Kneale, G., Firman, K. (1992) High-level expression of the cloned genes encoding the subunits of and intact DNA methyltransferase, M.EcoR124. *Gene* 112(1), 21–27.

59. Obarska, A., Blundell, A., Feder, F., Vejsadová, Š., Šišáková, E., Weiserová, M., Bujnicki, J. M., Firman, K. (2006) Structural model for the multisubunit Type IC restriction–modification DNA methyltransferase M.EcoR124I in complex with DNA. *Nucleic Acids Res.* 34(7), 1992–2005.

60. Kennaway, C. K., Taylor, J. E., Song, C. F., Potrzebowski, W., Nicholson, W., White, J. H., Swiderska, A., Obarska-Kosinska, A., Callow, P., Cooper, L. P., Roberts, G. A., Artero, J. B., Bujnicki, J. M., Trinick, J., Kneale, G. G., Dryden, D. T. (2012) Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Genes Dev.* 26(1), 92–104.

61. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J., Lesk, A. M. (2006) MUSTANG: a multiple structural alignment algorithm. *Proteins* 64(3), 559–574.

62. Krieger, E., Koraimann, G., Vriend, G. (2002) Increasing the precision of comparative models with YASARA NOVA – a self-parameterizing force field. *Proteins* 47(3), 393–402.

63. Bullas, L. R., Colson, C., Van Pel, A. (1976) DNA restriction and modification systems in Salmonella. SQ, a new system derived by recombination between the SB system of Salmonella typhimurium and the SP system of Salmonella potsdam. *J Gen Microbiol.* 95(1), 166–172.

64. Fuller-Pace, F. V., Murray, N. E. (1986) Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes. *Proc Natl Acad Sci U S A.* 83(24), 9368–9372.

65. Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A., Bickle, T. A. (1992) Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J.* 11(1), 233–240.

66. Weiserova, M., Firman, K. (1998) 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.

67. Weiserova, M., Dutta, C. F., Firman, K. (2000) 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(1), 301–310.

68. Dryden, D. T., Murray, N. E., Rao, D. N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.* 29(18), 3728–3741.

69. Kelleher, J. E., Daniel, A. S., Murray, N. E. (1991) Mutations that confer de novo activity upon a maintenance methyltransferase. *J Mol Biol.* 221(2), 431–440.

70. Cooper, L. P., Dryden, D. T. (1994) The domains of a type I DNA methyltransferase: Interactions and role in recognition of DNA methylation. *J Mol Biol.* 236(4), 1011–1021.

71. Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L., Ginalski, K. (2012) Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.* 40(15), 7016–7045.

72. Šišáková, E., Stanley, L. K., Weiserová, M., Szczelkun, M. D. (2008) A RecB-family nuclease motif in the Type I restriction endonuclease EcoR124I. *Nucleic Acids Res.* 36(12), 3939–3949.

73. Šišáková, E., Weiserová, M., Dekker, C., Seidel, R., Szczelkun, M. D. (2008) The Interrelationship of Helicase and Nuclease Domains during DNA Translocation by the Molecular Motor EcoR124I. *J Mol Biol.* 384(5), 1273–1286.

74. Sinha, D., Shamayeva, K., Ramasubramani, V., Řeha, D., Bialevich, V., Khabiri, M., Guzanová, A., Milbar, N., Weiserová, M., Csefalvay, E., Carey, J., Ettrich, R. (2014) Interdomain communication in the endonuclease/motor subunit of type I restriction-modification enzyme EcoR124I. *J Mol Model.* 20(7), 2334.

75. Szczelkun, M. D., Dillingham, M. S., Janscak, P., Firman, K., Halford, S. E. (1996) Repercussions of DNA tracking by the type IC restriction endonuclease EcoR124I on linear, circular and catenated substrates. *EMBO J.* 15(22), 6335–6347.

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

77. Davies, G. P., Martin, I., Sturrock, S. S., Cronshaw, A., Murray, N. E., Dryden, D. T. (1999) On the Structure and Operation of Type I DNA Restriction Enzymes. *J Mol Biol.* 290(2), 565–579.

78. Bialevich, V., Sinha, D., Shamayeva, K., Guzanova, A., Řeha, D., Csefalvay, E., Carey, J., Weiserova, M., Ettrich, R. H. (2017) The helical domain of the EcoR124I motor subunit participates in ATPase activity and dsDNA translocation. *PeerJ.* 5, e2887.

79. Dryden, D. TF. (2017) Structures of the type I DNA restriction enzymes. *Proc Natl Acad Sci U S A* 114(48), E10261–E10262.

80. Abadjieva, A., Patel, J., Webb, M., Zinkevich, V., Firman, K. (1993) A deletion mutant of the type IC restriction endonuclease EcoR1241 expressing a novel DNA specificity. *Nucleic Acids Res.* 21(19), 4435–4443.

81. Gubler, M., Bickle, TA. (1991) Increased protein flexibility leads to promiscuous protein-DNA interactions in type IC restriction-modification systems. *EMBO J.* 10, 951–957.

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

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

84. Mernagh, D. R., Taylor, I. A., Kneale, G. G. (1998) Interaction of the type I methyltransferase M.EcoR124I with modified DNA substrates: sequence discrimination and base flipping. *Biochem J.* 336(Pt 3), 719–725.

85. Mernagh, D. R., Kneale, G. G. (1996) 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.

86. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., Roberts, R. J. (1993) Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell.* 74(2), 299–307.

87. Bianco, P. R., Hurley, E. M. (2005) The type I restriction endonuclease EcoR124I, couples ATP hydrolysis to bidirectional DNA translocation. *J Mol Biol.* 352(4), 837–859.

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

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

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

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

92. Simons, M., Szczelkun, M. D. (2011) Recycling of protein subunits during DNA translocation and cleavage by Type I restriction-modification enzymes. *Nucleic Acids Res.* 39(17), 7656–7666.

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

94. Bianco, P. R., Xu, C., Chi, M. (2009) Type I restriction endonucleases are true catalytic enzymes. *Nucleic Acids Res.* 37(10), 3377–3390.

95. Jindrova, E., Schmid-Nuoffer, S., Hamburger, F., Janscak, P., Bickle, T. A. (2005) On the DNA cleavage mechanism of Type I restriction enzymes. *Nucleic Acids Res.* 33(6), 1760–1766.

96. Dreier, J., MacWilliams, M. P., Bickle, T. A. (1996) DNA cleavage by the type IC restriction-modification enzyme EcoR124II. *J Mol Biol.* 264(4), 722–733.

97. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 40(W1), W597–W603.

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

99. Qi, D., Scholthof, K. B. (2008) A one-step PCR-based method for rapid and efficient site-directed fragment deletion, insertion, and substitution mutagenesis. *J. Virol. Methods* 149(1), 85–90.

100. Miesenböck, G., De Angelis, D. A., Rothman, J. E. (1998) Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* 394(6689), 192–195.

101. Kabsch, W. (2010) XDS. *Acta Cryst. D* 66, 125–132.

102. Evans, P. (2006) Scaling and assessment of data quality. *Acta Cryst. D* 62, 72–82.

103. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. (2011) Overview of the CCP4 suite and current developments. *Acta Cryst. D* 67(4), 235–242.

104. Matthews, B. W. (1968) Solvent content of protein crystals. *J Mol Biol.* 33(2), 491–497.

105. Vagin, A., Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* 30, 1022–1025.

106. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., Read, R. J. (2007) Phaser Crystallographic Software. *J. Appl. Cryst.* 40, 658–674.

107. Cowtan, K. (2008) Fitting molecular fragments into electron density. *Acta Cryst. D* 64, 83–89.

108. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst. D* 67, 355–367.

109. Emsley, P., Lohkamp, B., Scott, W. G., Cowtand, K. (2010) Features and development of Coot. *Acta Cryst. D* 66(4), 486–501.

110. Yanisch-Perron, C., Vieira, J., Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* 33(1), 103–119.

111. Hubacek, J., Glover, S. W. (1970) Complementation analysis of temperature-sensitive host specificity mutations in Escherichia coli. *J Mol Biol.* 50(1), 111–127.

112. Asakura, Y., Kobayashi, I. (2009) From damaged genome to cell surface: transcriptome changes during bacterial cell death triggered by loss of a restriction–modification gene complex. *Nucleic Acids Res.* 37(9), 3021–3031.

pavel.grinkevich@gmail.com

Structure and Function of the C-terminal Domain of the HsdR Subunit from the Type I Restriction-Modification System EcoR124
Ph.D. Thesis Series, 2018, No. 10

Printed in the Czech Republic by Typodesign
Edition of 20 copies

University of South Bohemia in České Budějovice
Faculty of Science
Branišovská 1760
CZ-37005 České Budějovice, Czech Republic

Phone: +420 387 776 201
www.prf.jcu.cz, e-mail: sekret-fpr@prf.jcu.cz