University of South Bohemia, České Budějovice

Faculty of Science

Inter and intra domain interactions in the motor subunit of EcoR124I: A computational study

Ph.D. Thesis

Dhiraj Sinha, M.Sc.

Supervisor: Prof. RNDr Rüdiger H Ettrich, Ph.D.

Faculty of Science, University of South Bohemia,
and
Center of Nanobiology & Structural Biology
Institute of Microbiology
Academy of the Sciences of the Czech Republic

České Budějovice 2016

This thesis should be cited as:

Sinha D., 2016: Inter and intra domain interactions in the motor subunit of EcoR124I: A computational study. PhD Thesis, Series No. 15. University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic 164 pp.

Annotation

EcoR124I is a Type I restriction–modification (R–M) enzyme and as such forms multifunctional pentameric complexes with DNA cleavage and ATP-dependent DNA translocation activities located on the motor subunit HsdR. When non-methylated invading DNA is recognized by the complex, two HsdR endonuclease/motor subunits start to translocate dsDNA without strand separation activity up to thousands base pairs towards the stationary enzyme while consuming ~1 molecule of ATP per base pair advanced. Whenever translocation is stalled the HsdR subunits cleave the dsDNA nonspecifically far from recognition site. The X-ray crystal structure of HsdR of EcoR124I bound to ATP gave a first insight of structural/functional correlation in the HsdR subunit. The four domains within the subunit were found to be in a square planer arrangement. Computational modeling including molecular dynamics in combination with crystallography, point mutations, *in vivo* and *in vitro* assays reveals how interactions between these four domains contribute to ATP-dependent DNA translocation, DNA cleavage or inter-domain communication between the translocase and endonuclease activities.

Declaration

I hereby declare that I have worked on my doctoral dissertation thesis independently and used only the sources listed in the bibliography. I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of doctoral dissertation thesis, in full form resulting from deletion of indicated parts to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages. Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defense in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

Date: 09/Nov/2016 Signature: DHIRAJ SINHA

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

Financial Support:

PhD fellowship (2010-2014) from the Czech Science Foundation (project number GACR P207/12/2323), and the Grant Agency of the University of South Bohemia (grant no. 170/2010/P).

Dedicated to Dadi & Baba

Acknowledgement

First of all I would like to express my special appreciation and gratitude to my supervisor Prof. Rudiger Ettrich for his tremendous academic support and constant encouragement. It is his complete belief and trust in me that made possible to reach up to thesis defense. I always learned a lot, not only from his scientific discussion but his humble nature that inspired a lot and he helped me evolve both academically and personally.

Next I would like to thank almighty Shri Saccha Baba for his divine blessings always guiding me to walk on right path and my spiritual guru late Shri Vishnu Brahmchari for all his support to begin this journey. I would always remain thankful to my spiritual mentor Shri Hari Narain Pathak for all his efforts in putting my life back on right track.

I gratefully acknowledge Dr. Babak Minofar for introducing me to the group. Profound gratitude goes to Dr. David Reha for his effort in analyzing the simulation data, without his support I would not be able to expect any of my results. I am also thankful to all the computational lab members Natalia Kulik and Zofie Sovova, Sudhir and from experimental lab Pradeep, Pavel, Deepika, Katya and Vitali.

My sincere thanks also go to Vasilina Zayats, Anamika Mishra and Kumud Bandhu Mishra for all the support in the beginning of my studies. Special thanks to my friend Mr. Saurabh Kumar Pandey who enabled me to think differently and creating healthy environment in the lab.

I would like to express deep respect to my mentor Prof. Ashok Kumar Gupta who introduced me to the world of bioinformatics and established a new dimension to my academic career.

I do not have words to express my gratitude to my parents for their constant blessings and support. Nothing is comparable for all of the sacrifices that they've made. Their prayers for me helped me sustain thus far. I am really blessed to be the member of Sinha family headed by Dr Suman Kumar and always cherished our togetherness. I am also thankful to my family in-laws and happy to be part of new family as well.

This acknowledgement would be incomplete without the name of my beloved wife Mudra, who joined me in the most crucial phase of my PhD and made my life complete. Without her I could not have imagined any day of my life in every means.

List of publications and author's contribution:

The thesis is based on the following papers (listed chronologically).

[I] D Sinha*, K Shamayeva*, V Ramasubramani, D Reha, V Bialevich, M Khabiri, N Milbar, A Guzanova, M Weiserova, E Csefalvay, J Carey, R Ettrich (2014) Interdomain communication in the endonuclease/motor subunit of Type I restriction–modification enzyme EcoR124I. *Journal of Molecular Modeling 20(7):2334* [IF:1.864].

*(**DS** conceived the idea jointly with RE, performed MD simulations and analysis is done with DR and participated in writing manuscript).*

[II] David Reha, Balasubramanian Harish, **Dhiraj Sinha**, Zdenek Kukacka, James McSally, Olga Ettrichova, Petr Novak, Jannette Carey, Rüdiger Ettrich (2014) Molecular dynamics comparison of E. coli WrbA apoprotein and holoprotein. *Journal of Molecular Modeling* 20(9):2400 [IF:1.864]

***(DS** performed a significant part of the MD simulations*).

[III] Degtjarik Oksana, Brynda Jiri, Ettrichova Olga,Michal Kuty, **Sinha Dhiraj,** Kuta Smatanova Ivana, Carey Jannette, Ettrich Rudiger, Reha, David (2015) Quantum Calculations Using a New Crystal Structure Indicate effective Electron Transfer Between FMN and Benzoquinone in *E.coli* WrbA. *Journal of Physical Chemistry B* 120 (22):4867–4877 [1F:3.187]

(***DS** performed a significant part of the MD simulations*).

[IV] Bialevich V*, **Sinha D***, Shamayeva K, Reha D, Guzanova A, Csefalvay E, Carey J, Weiserova M and Ettrich R. The helical domain of the motor subunit of EcoR124I participates in ATPase activity and dsDNA translocation [accepted in PeerJ].

***(DS** is first author on the computational part, performed all MD simulations, did the computational analysis and participated in writing the manuscript).*

[V] Sinha D*, Bialevich V*, Shamayeva K, Reha D, Guzanova A, Sisakova A, Csefalvay E, Krejci L, Carey J, Weiserova M and Ettrich R The role of motif III and its extended region in positioning the two helicase domains in the motor subunit of the restriction–modification system EcoR124I [manuscript in preparation].

***(DS** is first author on the computational part, performed MD simulations and analysis and participated in writing the manuscript*).

[VI] Guzanova A, **Sinha D**, Pal K Sudhir, Ettrich R and Weiserova M Involvement of the conservative domains of the DNA recognition subunit HsdS in assembly of functional EcoR124I endonuclease [manuscript in preparation].

***(DS** contributed the computational part of the manuscript, performed MD simulations and participated in preparing figures*).

(*****Contributed equally)

Prof. RNDr. Rudiger H. Ettrich, Ph.D.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AdoMet	S-Adenosyl Methionine
AFM	Atomic Force Microscopy
ADP	Adenosine Diphosphate
ATP	Adenosine Triphosphate
DNA	Deoxyribonucleic Acid
dsDNA	Double-stranded DNA
EM	Energy Minimization
HJ	Holliday Junction
LINCS	Linear Constraints Solver Algorithm
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
PDB	Protein Data Bank
PME	Particle Mesh Ewald
QM/MM	Quantum Mechanics/Molecular Mechanics
REnase	Restriction Endonuclease
RM	Restriction Modification

RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuations
RNA	Ribonucleic Acid
SAD	Single-wavelength Anomalous Dispersion
SANS	Small Angle Neutron Scattering
SAXS	Small Angle X-ray scattering
SeMet	L-Selenomethionine
ssDNA	Single-stranded Deoxyribonucleic Acid
TBS	Triplex Binding Site
TFO	Triplex Forming Oligonucleotide
TRD	Target Recognition Domain
UV	Ultraviolet
VMD	Visual Molecular Dynamics
WT	Wild Type

Motivation

Bacterial diversity and prevalence across the prokaryotic kingdom depends upon its robust immune response. Bacteria have evolved to protect themselves from invading genome by nucleotide triphosphate-dependent *restriction*–modification systems. They comprise of two contrasting enzymatic activities, one restriction endonuclease (REase) and other methyltransferase (Mtase). The REase cleaves foreign unmodified DNA while Mtase activity ensures discrimination between self and non-self DNA by transferring methyl groups to the same specific DNA sequence within the host genome. Based upon the subunit structures and cofactors requirement, R–M enzymes have been classified into three main groups and EcoR124I belongs to Type I R–M system.

Type I R–M system are encoded by three genes, named *hsd* for host specificity determinant: *hsdR* encodes the restriction R subunit, *hsdM* the modification (M) subunit and *hsdS* the recognition (S for specificity) subunit. Based on genetic complementation, immunological cross reactivity and DNA hybridization, Type I enzymes can be grouped into three different families named IA, IB, and IC (Murray, Gough *et al.* 1982,Wilson and Murray 1991, Bickle and Kruger 1993). EcoR12I from *E.coli* is Type IC, hetero-pentameric complex with the stoichiometry of R_2:M_2:S_1 (Dryden, Cooper *et al.* 1997, Janscak and Bickle 1998). The S subunit recognizes an asymmetric sequence GAAnnnnnnRTCG where n is any base and R is purine (Price, Shepherd *et al.* 1987). The M_2S_1 component of this complex mediates its binding to the recognition sequence and can also function independently as a DNA Mtase (Taylor, Patel *et al.* 1992). The hsdR subunit is essential for endonuclease and

ATP dependent DNA translocation. A peculiarity of Type I R–M system is that they cut DNA in a nonspecific manner up to several thousand base pairs away from the recognition sequence. The unmethylated invading DNA is recognized by the complex, two HsdR endonuclease/motor subunit start to translocate dsDNA without strand separation activity. Translocation of up to thousand base pairs occurs towards the stationary enzyme by consuming ~1 ATP molecule of ATP per base pair advancement. During this activity long DNA loops are formed (van Noort, van der Heijden *et al.* 2004). Finally, when translocation is blocked by a stalling event, the HsdR subunit switches its conformation to a cleavage competent state and cleaves dsDNA nonspecifically.

In order to achieve better understanding of ATP-dependent DNA translocation and nuclease activity the first crystal structure of R subunit/motor subunit in complex with ATP has been reported in 2009(Lapkouski, Panjikar *et al.* 2009). The crystal structure shows a square planner arrangement of four distinct globular domains. The N-terminal endonuclease domain contains three catalytic residue (Asp151,Glu165 and Lys167) involved in cleavage activity. By fold and catalytic residues this domain is similar to the PD-(E/D)xK superfamily of Type II endonucleases and includes also a characterizedQ179xxxY183 motif as RecB nucleases (Sisakova, Stanley *et al.* 2008). The C-terminal helical domain is implicated in binding of HsdR to the HsdS−HsdM complex. In between endonuclease and helical domains two RecA-like helicase domains contain the seven conserved amino acids sequence motifs typical for helicase superfamily II that is relevant for ATP-dependent DNA translocation (Gorbalenya and Koonin

1991, Murray, Daniel *et al.* 1993). The variant DEAD box motifs are found in typical location to bind one ATP molecule bound in the cleft formed between two helicases.

In the crystal structure of the HsdR subunit ATP is unexpectedly in contact with Lys220 on the endonuclease domain. This unusual contact is not present in other helicases or translocases because of the absence of additional domains in solved structures. This is either due to the fact that resolved other structures represent just a helicase/translocase, or that within a larger complex the helicase domains form a subunit of its own without being covalently linked to additional domains. The additional interaction of endonuclease domain to the ATP molecule via Lys220 suggests a potential involvement of this residue in signal transfer from the translocase via the ATP binding pocket to the ~20A distant catalytic site through a so-called 180 loop located on the surface of the endonuclease domain. The equivalent position of Lys220 in HsdR subunit of Ecor124I is fully conserved throughout the Type IC family. The role of the positively charge lysine residue in contacting the ATP molecule is revealed through mutagenesis, crystallization and activity test and suggesting the contact of Lys220 of endonuclease domain with ATP molecule triggers the conformation from switching translocase competent conformation to cleavage competent conformation of the complex (Csefalvay, Lapkouski *et al.* 2015).

Although the 180 loop (residues182-189) is not resolved in the WT crystal structure, probably due to high flexibility, in the crystal structure of another mutant, Lys220Ala, the loop including the full motif QxxxY is resolved and placed above the endonuclease active site. A crystal

structure unfortunately represents only a single snapshot of one conformation and cannot give the full answer for inter-domain communication, and therefore my work started with the hypothesis about the role of 180 loop in signal transfer from ATP binding pocket to catalytic site through contact of Lys220-ATP in the motor subunit of EcoR124I.

It has been observed that in SF2 translocases the helicase2 domain actively participates in ATP dependent DNA translocation and that rotation of helicase 2 domain with respect to helicase1 domain occurs (Lewis, Dürr *et al.* 2008). Helicases contains sequence motifs known as DExx box helicases for ATP hydrolysis and ATP dependent DNA translocation.(Caruthers and McKay 2002). Particular conformations of the helicase 2 domain are essentially required to complete the translocation cycle and the conformation of helicase2 thus plays a key role to achieve the next step. Various crystal structures of helicases of super family 2 are reported in different conformations. In Rad54, a 180 degree of rotation of helicase 2 with respect to helicase 1 occurs during the binding of DNA (Dürr, Korner *et al.* 2005). In the crystal structure of HsdR of EcoR124I are specific inter-domain contacts at the helical-helicas2 interface. I was interested in revealing the role of these inter-domain interactions with the hypothesis that these interactions might play a major role in stabilizing the conformation of domains during a particular stage of the translocation cycle and thus point mutations to alter these interactions would affect the activity of the whole complex.

Recently a role for an extended motif III in dsDNA binding which is likely linked to ATPase activity was reported for Rad54 (Zhang, Janke *et*

al. 2013). I focused on identifying the corresponding residues of the extended motif III in our crystal structure of HsdR (2W00) of EcoR124I and their effect on the activity in DNA binding as well as ATP dependent DNA translocation and cleavage activity by mutational studies to explain role of these residue either from extended motif III or even from motif III in the functionality of the complex.

In one mutant crystal structure (E165H) of HsdR an alternative conformation of 180 loop was found in which R182 is neither in contact with 220 loop nor does it is in contact with D881. R182 instead is in contact with D855 and makes a salt bridge contact with the helical domain. With the help of molecular dynamics simulation alternative conformations were explored how the 180 loop interacts with D855 and if it might have a similar function as D881 in WT. Also conformational changes of endonuclease domain were studied in detailed and compared to the ones reported earlier.

On the methyltransferase, the HsdS subunit plays a key role in DNA recognition and subunit assembly of the pentameric complex. Our collaborators from Prague reported residues in the hinge region, between the central conserved and highly variable regions of the target recognition domains (TRD) to play an important role with respect to interactions within the methyltransferase, with the HsdM subunit, but also in assembly of the full complex with the HsdR subunits. (Janscak, Weiserova *et al.* 2000, Weiserova, Dutta *et al.* 2000). Recent work shows K184 and K384 which are the part of extended conserved region to affect assembly as well as activity of the complex. This gained my

interest in finding structural changes associated with these point mutations which might affect the subunit assembly and activity as well.

1. Introduction

Species have developed multilayer defense mechanisms to protect themselves for survival in the nature. In mammals various kinds of immune response like cell-mediated or adaptive immune response are observed. Viruses are combated by apoptosis of infected cells, such as natural killer cells destroy the non-competent infected cells or by producing antibody against any antigen. Sometime they also respond via RNA interference mechanism by inhibiting the gene expression. On the other hand plants mostly control foreign attack by programmed cell death through hypersensitive response. Whereas insects have intracellular protein-mediated antiviral defense mechanism, archaea and bacteria also contain numerous lines of defense against their viruses and phages such as by receptor masking, blocking DNA injection (Sie), *restriction*–modification (R–M) and R–M-like systems, abortive infection (Abi), and recently identified clustered regularly interspaced short palindromic repeats (CRISPR-associated) systems.

1.1 Bacterial Defense System

Bacterial diversity and prevalence across the prokaryotic kingdom depends upon the success of its defense system against the invading genome and its evolution depends on the ability to recognize and distinguish incoming foreign DNA and self DNA. Different types of defense system evolved very rapidly in prokaryotes and there is antagonistic co-evolution that exists between host and parasite (Gomez and Buckling 2011). Prevention of the virus from attaching to the bacteria is the first line of defense.

1.1.1 Preventing viral attachment

This process is driven by the expression of polysaccharide to mask the receptor on the cell surface or by mutating the receptor molecule of the host that block phage adsorption. This is the most common encountered mechanism to escape the phage predation but phages counteract it by mutation leading to recognition of different receptors (Hyman and Abedon 2010).

1.1.2 Blocking DNA injection

This is the second line of defense in which plasmid or prophage encode for the protein in the proximity of cell surface to resist the penetration of viral genome into the cytoplasm. The whole process is also known as Sie that gives a selective advantage to the prophage carrying bacterial population over non prophage containing bacteria. At some stage the prophage starts to multiply, proliferate and act as weapon to kill the prophage free bacteria This is also an example of mutual relationship between Sie encoding prophage and host bacteria (Roossinck 2011).

1.1.3 Abortive infection

It is an apoptosis like mechanism as in eukaryotes where the prokaryotic cells induce cell death upon viral infection to protect other cells of the same species. In *E.coli* several Abi systems are found either with code for single protein such as LitA, PrrC, PifA or multiple proteins (Labrie, Samson *et al.* 2010). Recently, a new class of Abi systems was reported by Fineran and coworkers called Type III TA systems where the proteins act as Toxin and RNA acts as Antitoxin (Blower, Pei *et al.* 2011).

1.1.4 Adaptive Immunity

Also known as the CRISPR/CAS system, it is the most recent immune response discovered in bacteria and archaea. In this system there are some short repetitive base sequences and short spacer DNA that originated from the end of the previous exposed virus. It counters the foreign genetic material similar to RNAi in eukaryotic organism. The whole mechanism of CRISPR/CAS passes through several stages: adaptation, expression and interference (Bhaya, Davison *et al.* 2011). In the adaptation stage resistance is acquired by integration of a new spacer sequence in the CRISPR array followed by CAS gene expression and translation.

1.1.5 Digestion of Non-self DNA

Besides the other defense systems, intracellular mechanisms act directly on viral DNA to neutralize the invader. R–M systems are part of a wide range of prokaryotic immune responses that targets DNA. It occurs exclusively in unicellular organisms, mainly in bacteria. None have been found in eukaryotes, although certain chlorella viruses encode them (Rohozinski, Girton *et al.* 1989). They have been found roughly in one quarter of bacteria examined to date, the remaining three quarters might lack R–M systems. The phenomenon of R–M system was first reported by Luria and Human in 1952 when they tried to explain host-control variation in bacterial viruses. They observed the efficiency with which phage infected new bacterial hosts which depends on the host on which they previously grew. Phage that propagated efficiently on one bacterial strain could lose the ability if grown for even a single cycle on a different strain (Luria 1953) (Bertani and Weigle 1953). The term “restriction”

comes from the observation that some strains of *E.coli* greatly reduced the ability of bacteriophages to form plaques as observed by Werner Arber and Daisy Dussoixin 1960(Dussoix and Arber 1962). However, bacteriophages that have escaped restriction and form plaques have been "modified" and are now resistant to restriction if re-plated on the same host. This resistance is specific to the particular R–M system and is lost when the bacteriophages are grown in a different host strain. Arber found that the phenomenon had two opposing aspects, restriction and modification and proposed that each was catalyzed by an enzyme that would in one case cut the DNA and in the other protect it. He demonstrated that protection involved transfer of methyl group from S-adenosyl L-methionine (Adomet) (Kuhnlein and Arber 1972). Later in 1968 Linn and Arber were successfully able to purify classical R–M system EcoKI and EcoBI (Linn and Arber 1968).

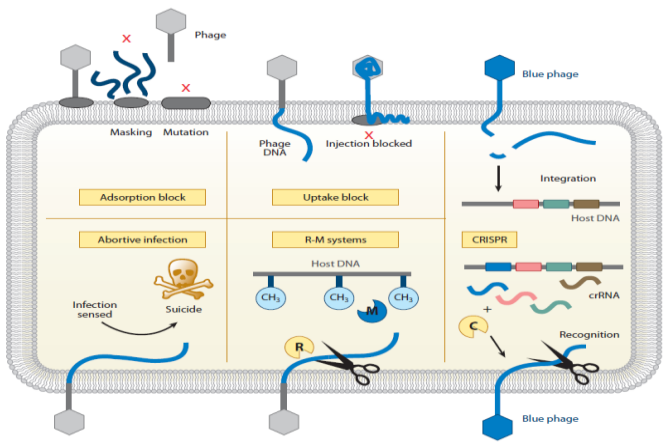


Fig.1.1: Schematic overview of prokaryotic defense mechanism. Several independent defense mechanisms present in the bacterial cell including blocking of phage, restriction/modification system or clustered regularly interspaced short palindromic repeats. (Figure adapted from Edze R *et al.* 2012).

In 1970 Hamilton Smith confirmed Arber's hypothesis (Smith and Wilcox 1970) and found out that all DNA fragments generated by this enzyme had the same nucleotides at their ends, indicating cleavage within short symmetrical sequences. This was a remarkable milestone in molecular biology and later named as Type II restriction enzymes (Kelly and Smith 1970). Daniel Nathans used Smith's findings and generated a restriction map for the small animal virus further suggested several other important uses for restriction enzymes (Danna and Nathans 1971). The use of restriction endonuclease along with Agarose gel was first reported by Sharp and coworkers (Sharp, Sugden *et al.* 1973).

In 1973 Smith and Nathans (Smith and Nathans 1973)suggested the standard nomenclature for REase and DNA Mtase. Werner Arber, Hamilton Smith, and Daniel Nathans jointly shared the Nobel Prize for Physiology or Medicine in 1978 for "*The discovery of restriction enzymes and their application to problems of molecular genetics*". Smith and Nathans proposed that enzyme name should begin with a three letter acronym where the first letter belongs to the first letter of genus from which the enzyme was isolated and the next two letters was the first two letters of the species name. Extra letters or numbers could be added to indicate individual strain or serotypes. Later more enzymes mostly Type II (3500) have been characterized often from different genera and species with names whose three letter acronyms would be identical, then Dr Richard Roberts (Roberts, Belfort *et al.* 2003) proposed the most systemic approach which incorporates current knowledge about the complexities of these enzymes. He also maintains the most comprehensive database on restriction modification system. Currently

complete sequenced genomes about 5000 R–M system are available through REBASE (Roberts, Vincze *et al.* 2015) in several different formats out of which 112 Type-I, 4060 Type-II, 22 Type-III and 19 Type IV are characterized till date.

1.2. Classification

The first classification attempt was made by Boyer (Boyer 1971) who divided these enzyme into two main group Type I or ATP dependent enzymes and Type II or ATP independent enzymes. However, the EcoPI, EcoP15I and HindIII system which are all ATP dependent, showed considerable difference from other members of Type I class which led the formal definition of Type III class of R–M systems (Kauc 1992). Therefore, the whole system is divided into three main groups Type I, Type II and Type III depending on the subunit composition, cofactor requirements and the nature of the reaction substrates and products. In addition, a fourth type has been identified that consist of restriction enzymes that cleave modified DNA alone (Roberts, Belfort *et al.* 2003). Additionally, new systems are being discovered that do not fit readily into this classification.

Table 1. Classification of R–M systems (adapted from Sistla and Rao, 2004; Loenen *et al.*, 2014).

Characteristic	Type I	Type II	Type III	Type IV
Subunits	Three different	Two identical	Two different	Two different
Enzyme activity	Endonuclease, methylation, ATPase	Endonuclease or methylation	Endonuclease, methylation, ATPase	Endonuclease
DNA translocation	Yes	No	Yes	Yes
Cofactor for DNA cleavage	ATP, Mg^{2+} (AdoMet)	Mg^{2+}	ATP, Mg^{2+}, AdoMet	GTPase, Mg^{2+}
Cleavage site	Random, remote from recognition site	At or near recognition site	25-27 bp from recognition site	Between methylated bases at multiple positions
Recognition sequence	Asymmetric, bipartite	Symmetric	Asymmetric	Bipartite, methylated
Methylation	Mg^{2+}, AdoMet	AdoMet	Mg^{2+}, AdoMet	No

1.2.1 Type I R–M systems

Type I R–M systems are the most complex of all known systems with both restriction and modification activities carried out by the same enzymes (Davies, Martin *et al.* 1999, Rao, Saha *et al.* 2000). They are composed of three subunits: the motor subunit HsdR responsible for DNA translocation and cleavage, the modification subunit HsdM responsible for methylation of hemimethylated host DNA and the specificity subunit HsdS responsible for DNA recognition, most likely in the stoichiometry of $R_2M_2S_1$ (Dryden, Murray *et al.* 2001)**.** These systems recognize asymmetric bipartite sequences and cleave DNA 1000-7000bp away from the recognition site in the presence of divalent

cation Mg^{2+} and use ATP as a source of energy. If the pentameric complex binds to a site that is methylated in one strand, the methyltransferase gets stimulated and methylates the unmodified strand using Adomet and ATP.

1.2.2 Type II R–M systems

Type II R–M systems are the simplest modification-M systems in which the methylation and cleavage functions are carried out by distinct proteins encoded by independent genes: one for restriction and another for modification which act independently as well. These enzymes recognize short palindromic sequences of 4-8 bp in length and cleave within the recognition sequence and requires simple co-factor Mg^{2+} for restriction and Adomet for modification activity. They are classified on the basis of cleavage position and recognition site (Kessler and Manta 1990). The high specificity of DNA sequence recognition and cleavage by Type II nucleases has allowed extensive application of these enzymes as tools for dissection of DNA and in the construction of recombinant genes.

1.2.3 Type III R–M systems

Type III R–M systems are hetero-oligomeric, multifunctional proteins composed of Mod subunit responsible for site-specific methylation of the DNA and the Res subunit that is able, after formation of a $(Mod)_2(Res)_2$ complex, to translocate and cleave DNA in an ATP-dependent fashion (Haberman 1974). The recognition site of the enzymes is usually 5-6 nucleotides long and asymmetric but not bipartite. The cleavage site for the restriction endonuclease is 25-27 bp downstream of the recognition

sequence, in contrast to the non-specific distance of thousands of bases for the Type I enzyme (Kauc and Piekarowicz 1978). The peculiar feature of these enzymes is that they methylate only one strand of the DNA, at the N6 position of adenosyl residues, thus newly replicated DNA always has one strand methylated which is sufficient to protect the self-DNA against restriction (Rao, Saha *et al.* 2000).

1.2.4 Type IV R–M systems

Type IV R–M systems are enzymes that contain both methylation and restriction functions on the same polypeptide and require Mg^{2+} and AdoMet for cleavage. DNA cleavage occurs at 14-16 bp away from the recognition sequence. These enzymes have been suggested to be the evolutionary link between the Type III and Type IIS enzymes (Janulaitis, Vaisvila *et al.* 1992). The ability of Type IV enzymes to distinguish between C,m5C,hm5C and other molecular variations of cytosine implicates that these are useful tools for studies of epigenetic phenomenon (Ordway, Bedell *et al.* 2006). The commercially available enzyme McrBC has been used for the study of such modification patterns (Irizarry, Ladd-Acosta *et al.* 2008).

1.3 Additional role of R–M systems

R–M systems are highly sequence specific and show low sequence homology. Sometimes in many bacterial species multiple R–M systems are found which suggest that they are not only involved in the cellular defense but also they have very diverse set of roles.

Kobayashi tried to explain the behavior of R–M system as selfish gene loci and the effect of their mobility on the genome (Kobayashi 2001).

This is according to the selfish gene hypothesis which is based on the observation of the Type II enzyme PaeR7. The hypothesis states that after a number of cell divisions, when the methyltransferase level is not high enough to protect the genome, the cells ultimately die because of lack of plasmid. The role of the selfish gene behavior of R–M systems thus would be maintaining the methylation status of the genome (Kusano, Naito *et al.* 1995). It was observed that R–M system are often linked to mobile elements or acquired through horizontal gene transfer and genomic R–M systems may also play a role in stabilizing the host chromosome. The importance of having the multiple R–M system can be partly explained by their role in the stabilization of genomic islands. Several previous works indicate the role of R–M system in recombination (Chang and Cohen 1977), nutrition and in genetic diversity, genomic integrity, regulation of genomic flux, stabilization genomic islands or in maintenance of the methylation pattern. Thus, bacterial cells seem to utilize R–M systems in various biological processes to increase the relative fitness of the population.

1.4 Overview of TypeI R–M systems—EcoR124I

Type Is are known as the most complex systems (Redaschi and Bickle 1996), composed of three subunits: HsdS (50 kDa), HsdM (50-60 kDa) and HsdR (140 kDa) (Murray 2000), (Rao, Saha *et al.* 2000) and the resulting complex exhibits endonuclease activity, methyltransferase activity and additionally translocase activity. These enzymes are categorized into four families, IA, IB, IC and ID on the basis of genetic complementation, DNA hybridization, sequence homology and antigenic cross reactivity (Barcus, Titheradge *et al.* 1995). Typical members of

Type IA are EcoBI and EcoKI, EcoAI for Type IB, EcoR124I and EcoR124II for Type IC and StyBLI for ID. The *hsd* genes of Type IA and IB are chromosomally encoded while IC systems are generally coded by the plasmid (Firman, Creasey *et al.* 1983). The KpnBI system which does not have enough resemblance with the pre-existing groups of Type I R–M systems is sometimes also classified as IE family (Chin, Valinluck *et al.* 2004).

The Type IC EcoR124I R–M system was first purified by Hughes (Hughes 1977) from the plasmid R124 (Meynell and Datta 1966) and able to show that the enzyme required ATP as cofactor and DNA cleavage activity was at least stimulated by S-adenosyl methionine/AdoMet while the genetic evidence is given earlier by Bannister and Glover (Bannister and Glover 1968). Firman and group showed major physical difference between R124 and R124/3 and described the complex DNA rearrangement between the two plasmids (Firman, Creasey *et al.* 1983).

Bickle's group reported the difference between recognition sites of EcoR124 and EcoR124/3. They found that the two specificities only differ by inclusion of one extra base in the spacer region. The recognition site is $GAAn_6RTCG$ in EcoR124I while $GAAn_7RTCG$ in EcoR124/3 (Price, Shepherd *et al.* 1987). They also confirmed that there was no structural difference between the two but they are antigenic ally different from other Type I R–M system (Price, Pripfl *et al.* 1987).

A typical Type I R–M system is composed of three subunit encoded by three closely linked genes *hsdM*, *hsdS* and *hsdR* in which HsdM and HsdS transcribed by same promoter but HsdR have its own independent

promoter (Murray 2000). The *hsdR* gene can operate functionally only along with *hsdS* and *hsdM* genes (Glover 1970). The dimeric HsdM subunit together with monomeric HsdS subunit forms the trimeric complex methyltransferase which is able to methylate unmodified DNA in the presence of cofactor AdoMet. Methyltransferase of different species varies in size and structure. The size of each HsdM subunit is 58 kDa in EcoR1241 and that of HsdS subunit is46 kDa; thus the molecular weight of the whole trimeric methyltransferase complex is 162 kDa (Taylor, Davis *et al.* 1994).

1.4.1 Modification subunit HsdM

Methylation of host DNA occur either in cytosine at C5 position or N4 position, or adenine at the N6 position from donor Adomet by the methyltransferase (Wilson and Murray 1991). Methylation only occurs when HsdM subunit is integrated in the complex with a stoichiometry of $HsdM_2$:$HsdS_1$. Although the sequence of the HsdM subunit is poorly conserved it contains two characteristic motifs for binding of AdoMet and methylation catalysis. Hattman and Ahmad postulated the role of motif I [(D/E/S)X,F,X,G,X,G] in binding of cofactor AdoMet and motif IV [(N/D/S)P,P(Y/F)] which plays a key role in catalysis (Hattman, Wilkinson *et al.* 1985, Ahmad and Rao 1996). Willcock and Dryden showed the effect of mutation in conserved motif I and IV in binding with AdoMet and catalysis (Willcock, Dryden *et al.* 1994). In 2006, Obarska proposed the model of catalytic domain of HsdM of EcoR124I using Frankenstein monster (Kosinski, Gajda *et al.* 2005) approach with domains of N- and C-terminal regions modeled based upon the crystal structure of HsdM of EcoKI (PDB ID: 2AR0) from *E.coli* as template.

However, the orientation of this domain still remains unknown (Obarska, Blundell *et al.* 2006). One other crystal structure of HsdM from *Bacteroides thetaiotaomicron*: M.BthVORF4518P at 2.2 Å is also reported. Methylation is probably achieved via base-flipping mechanism in which the target base is rotated 180 degree out of the DNA helix into the catalytic site of the enzyme (Mernagh, Taylor *et al.* 1998).

1.4.2 Specificity subunit HsdS

The HsdS subunit recognizes the target DNA sequence and serves as a core subunit for binding of other subunits. It contains duplicated organization of the two TRDs in tandem. The N-terminal TRD recognizes the 5′ part of the bipartite target sequence and the C-terminal TRD recognizes the 3′ part of the target (Fuller-Pace and Murray 1986, Gubler, Braguglia *et al.* 1992). The TRD sequences of HsdS subunit are poorly conserved showing less than 30% identity unless they recognize the same target sequence. Sturrock and Dryden in 1997 reported that TRD had common tertiary structure with a DNA binding region similar in structure to that found in the crystal structure of the TRD of the type-II methyltransferase HhaI (Sturrock and Dryden 1997). The TRDs are separated by a short amino acid spacer region which is highly conserved between within a Type I family and less strong conserved between families. The similarity between different families in this region is mostly confined to a short sequence which suggested that HsdS subunit possesses ~2-fold rotational symmetry (Kneale 1994). The EcoR124I and EcoR124II members of the Type IC family differ only in the length of the non-specific spacer sequence by one nucleotide (Price, Lingner *et al.* 1989). Deletion of one of the repeats (four amino acids T-A-E-L) in

EcoR124II altered its specificity to that of EcoR124I. HsdS varies in solubility *in vitro* depending upon the Type I system, EcoKI is insoluble while EcoR124I is only soluble as fusion protein (Mernagh, Janscak *et al.* 1998). The HsdS subunit is completely insoluble unless it is co-expressed with HsdM subunit forming Mtase complex (Patel, Taylor *et al.* 1992). There was no direct evidence to identify which amino acid residues within each TRD interact with DNA until two crystal structure of HsdS subunit was reported in 2005. One S.MjaORF132P from *Methanocaldococcus jannaschii* solved at 2.4 Å (PDB ID: 1YF2)(Kim, DeGiovanni *et al.* 2005) and other S.MgeORF438P from *Mycoplasma genitalium* solved 2.3 Å (PDB ID: 1YDX) (Calisto, Pich *et al.* 2005) and in 2006 Obarska proposed the model of HsdS subunit of EcoR124I based on these crystal structures by Frankenstein's monster approach (Obarska, Blundell *et al.* 2006). These models allowed the identification of a number of functionally important residues that appear to be involved in DNA binding.

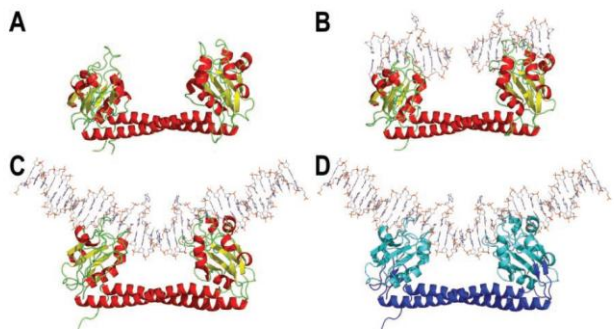


Fig.1.2: Predicted 3D structure of DNA binding subunit (HsdS) of EcoR1241. A: Crystal structure of HsdS (MjaXI) B: Model generated using Frankenstein's approach C: Energy minimized model of HsdS with DNA bound state as experimentally measured at 49° bend D: Mapping of the conserved S helix region (blue) and the TRDs (cyan) Figure adapted from (Obarska *et al.* 2006).

Classically, mutation studies in the central conserved region of HsdS subunit produced the (R^-,M^-) phenotype (Glover and Colson 1969, Hubacek and Glover 1970)but Weiserova reported also mutations in HsdS subunit of EcoR124I which gave surprisingly the phenotype of (r^-,m^+)(Weiserova and Firman 1998) and suggested that mutations in HsdS subunit alter the interaction of HsdR subunit with methyltransferase. Further *in vitro* experimental studies showed that the effect of these mutations was more complex and appear to alter the ability of the methyltransferase to undergo conformational changes required for DNA binding (Weiserova, Dutta *et al.* 2000). Several mutational analyses within the connecting region between TRDs and coil-coil structure confirmed its role as flexible hinge. These hinge residues helps in the movement of two TRDs induce the conformational changes associated not only with DNA binding but also bending of the enzyme upon DNA binding (Calisto, Pich *et al.* 2005, Kim, DeGiovanni *et al.* 2005, Su, Tock *et al.* 2005).

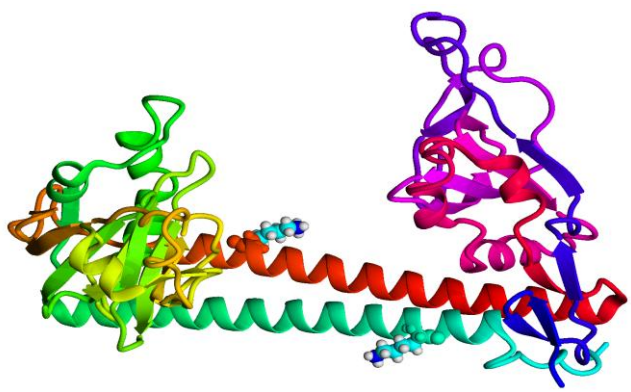


Fig.1.3: Homology model of HsdS subunit based on crystal structure of HsdS from *M. jannaschii* (PDB ID: 1YF2). The residues Lys 184 and Lys 384 are shown as ball representation. Figure prepared in Yasara.

1.4.3 Trimeric methyltransferase complex

Methylation activity is carried out by the trimetric methyltransferase, which is composed of two HsdM subunits and one HsdS subunit (Taylor, Patel *et al.* 1992,Dryden, Cooper *et al.* 1993, Janscak and Bickle 1998). Taylor *et al.* carried out the purification and biochemically characterized the EcoR124I methyltransferase complex. In EcoKI and EcoBI the methyltransferase dissociates into M_1S_1 and M_1 but the functional complex is always present as M_2S_1 (Dryden, Cooper *et al.* 1993). The dimeric M_1S_1complex exhibits weaker affinity for the EcoKI target sequence than the trimeric one and has slightly less ability to discriminate against other DNA sequence (Powell, Dryden *et al.* 1998). The functional form of methyltransferase catalyzes the transfer of methyl group from AdoMet to the N6 position of a specific adenine residue in each part of the recognition site, when the target base is rotated 180 degree out of the DNA helix towards the catalytic site of enzymes (Nagaraja, Shepherd *et al.* 1985),probably utilizing the cation-π interactions proposed for those Type II enzymes that methylate adenine at the N6 position or cytosine at the N4 position (Schluckebier, Labahn *et al.* 1998). DNA binding is enhanced in the presence of AdoMet and it probably plays a role in distinguishing between unmodified and modified DNA. Foot printing experiment shows that two HsdM subunits are located on either side of DNA helical axis rather than lying along the helical axis (Powell and Murray 1995). This experiment also determines that hsdM of EcoKI and EcoR124I cover approximately 25-30 bp in the major groove of DNA helix and the methylation state of the recognition sequence had no effect on the binding affinity. The affinity of enzyme

varies in the family and depends on the methylation state of DNA. EcoAI from IB family shows no affinity for hemimethylated DNA while in EcoR124I the catalysis activity of methylation is preferred in hemimethylated DNA rather than unmethylated DNA (Taylor, Watts *et al.* 1993).

DNA binding induces major conformational changes in TypeI methyltransferases and Taylor based on small angle X-ray scattering (SAXS) measurement proposed that binding of DNA causes a large rotation of the two HsdM subunit towards the DNA which is mediated by hinge bending domains in the HsdS subunit (Taylor, Davis *et al.* 1994). However, UV CD spectroscopy revealed that there is no significant change in the secondary structure of the subunits of the methyltransferase when its DNA bound. Conformational changes in methyltransferase leads to structural changes in DNA which involve local unwinding of the helix or increase in propeller twist of the base pairs (Johnson, Dahl *et al.* 1981). These local unwinding of DNA may be required for the catalytic residues of methyltransferase and the methyl group donor S-adenosyl methionine to gain proper access to the N6 amino group of adenine that is the target for the methylation by the enzyme. Callow (Callow, Sukhodub *et al.* 2007) first reported the low resolution structure of DNA methyltransferase M.AhdI using the small-angle neutron scattering (SANS) method with the stoichiometry M_2:S_2 but the properties are similar to Type I methyltransferase. Specific deuteration and contrast variation techniques revealed the location of individual subunits and proposed the extended regions of the HsdM subunit in Type I methyltransferases that are flexible and collapse around

the DNA to form a more globular structure of the methyltransferase-DNA complex.

Using bioinformatics methods Obarska predicted a structural model for the multi-subunit Type IC R–M EcoR124I methyltransferase in complex with DNA comprising the HsdM and HsdS subunits, based on the crystal structure of HsdS (MjaXIP), HsdS (MgeORF438P) and the M.TaqI-DNA complex with docking of domains guided by experimental data on protein-DNA interaction in EcoR124I (Taylor, Davis *et al.* 1994). This model of the EcoR124I complex has given a first structural context for sequence conservation between HsdS (EcoR124I) and related HsdS subunits in particular StySKI and also describes the location of a number of DNA binding mutations within the HsdS gene of EcoR124I (Obarska, Blundell *et al.* 2006).

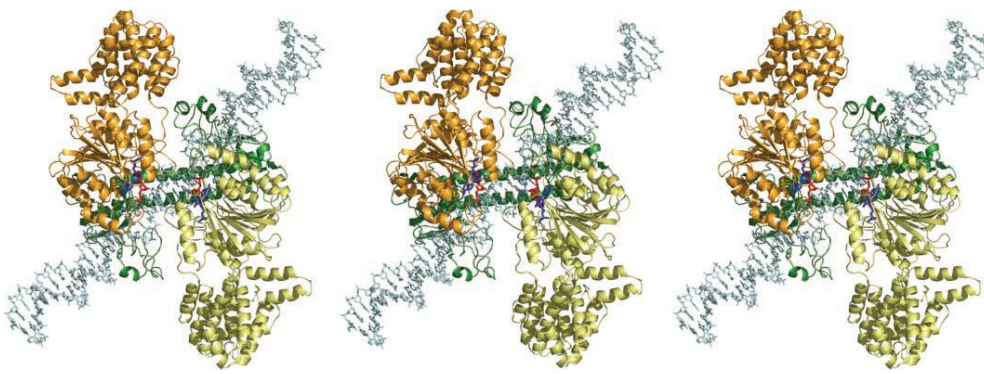


Fig.1.4: Computationally predicted structure of the EcoR124I methyltransferase and substrate. Two HsdM subunits are showed in yellow and orange. The HsdS subunit is shown in green. All three images are from different angles. Figure adapted from (Obarska *et al.* 2007).

Obarska and coworkers also mapped surface-modifiable lysine residues of the EcoR124I methyltransferase and validated with the experimental data of Taylor and coworkers. The solvent accessibility surface area of

lysine within the structural models of HsdS subunit and EcoR124I+DNA, were analyzed with the available model and it was possible to examine the location of the accessible lysines more closely.

Several other models of methyltransferase were also proposed (Kneale 1994, Dryden, Sturrock *et al.* 1995) but the overall crystal structure of the methyltransferase is still not known. Kennaway and coworkers presented a model at 35Å resolution determined by negative-staining electron microscopy of trimeric complex methyltransferase of EcoKI bound to ocr. The dimeric ocr protein from T7 phage resembles dsDNA highly by showing typical characteristics of dsDNA bend to the degree necessary for recognition by EcoKI. This model suggests dynamic opening and closing of protein driven by conserved coiled-coil region of HsdS subunit required to allow DNA binding (Kennaway,Obarska-Kosinska *et al.* 2009).

1.4.4 Motor subunit HsdR

The motor subunit is the biggest subunit of pentameric complex and possesses cleavage and translocation functions. HsdR subunit has a lower binding affinity to DNA and its binding affinity is independent of methylation state of DNA. Two HsdR subunits sequentially bind with the methyltransferase to form the complete functional pentameric restriction-modification complex of EcoR1241 and cleaves DNA distantly from the recognition site involving ATP-dependent dsDNA translocation (Horiuchi and Zinder 1972, Murray, Batten *et al.* 1973, Brammar, Murray *et al.* 1974). The binding affinity of the first HsdR is relatively much stronger than the one of the second subunit (Janscak, Dryden *et al.* 1998, Mernagh, Janscak *et al.* 1998). The complex with a

stoichiometry of R_1:M_2:S_1 shows ATPase activity but no cleavage activity was observed. Cleavage activity is only possible when the second HsdR subunit binds to the complex (Janscak and Bickle 1998). The R_2 complex shows a higher translocation rate than the R_1complex. Szezelkun and coworkers proposed a kinetic model of translocation and showed that the HsdR subunit is able to dissociate with the pentameric complex and starts translocation by resetting itself on the DNA (Szczelkun 2002).

At the amino acid sequence level, the HsdR subunit is poorly conserved and possesses 20-30% overall identity within the four representative families of Type I systems (Titheradge, Ternent *et al.* 1996). However, it contains several characteristic motifs positioned on different domains to perform all enzymatic functions. These motifs are organized into the modular structure of DNA binding and ATP binding helicase domains and of one endonuclease domain.

The helicase domain catalyzes DNA translocation using hydrolysis of ATP as source of energy (Matson, Bean *et al.* 1994). Based on several conserved motifs, helicases have been classified into six super families (Ilyina, Gorbalenya *et al.* 1992). Superfamily I (SFI) and Superfamily (SFII) generally participate in manipulation of DNA and RNA including splicing, translation or DNA degradation (Rocak and Linder 2004). They are found in monomeric forms while other helicases have only one RecA like domain with two to five conserved motifs making large hexameric rings and often acting on the replication fork (Donmez and Patel 2006). The HsdR subunit of EcoR124I is composed of seven conserved motifs called DEAD box motifs. This suggests that it belongs to SFII of DNA

and RNA helicases. These DEAD box motifs make a nice grove as a binding pocket between the two RecA-like domains, Helicase I and Helicase II to incorporate ATP. Some portions of these conserved motifs also form nucleic acid binding sites (Hall and Matson 1999).

Sequence analysis and mutational studies reveal the importance of these motifs in ATP hydrolysis and DNA translocation (Murray, Daniel *et al.* 1993, Davies, Powell *et al.* 1998, Davies, Martin *et al.* 1999). Several crystal structures have been reported for helicases of SFI and SFII in which the enzyme crystallized as a monomer consisting of two parallel domains, with a deep cleft between them (Delagoutte and von Hippel 2002). These adjacent domains are covalently linked with each other and each contains five beta strands surrounded by five alpha helices and are folded like RecA ATPases (Story and Steitz 1992, Caruthers, Johnson *et al.* 2000).

Motif I also known as Walker A and Motif II or Walker B is common among all helicases (Walker, Saraste *et al.* 1982, Gorbalenya, Koonin *et al.* 1988). These two motifs from the ATP binding site and are very crucial for ATPase activity. Mutation of residues in the conserved motif abolishes the ATPase activity (Rozen, Pelletier *et al.* 1989, Blum, Schmid *et al.* 1992). It was shown that they are involved in ATP binding by forming hydrogen bonds with the phosphates and by interacting with the coordinated Mg^{2+} ion (Caruthers and McKay 2002, Cordin, Tanner *et al.* 2004). Motif IA and IB in association with motif IV and motif V participate in nucleic acid binding (Rogers, Komar *et al.* 2002). Some studies also show that Motif IA participates in the structural rearrangement that occurs upon ATP binding and hydrolysis (Story, Li *et*

al. 2001, Zhao, Shen *et al.* 2004). Motif III is proposed to participate in linking ATPase and helicase activities and as in other SFI Motif III has been shown not to be part of the ATP binding pocket, as can be seen from the crystal structure of HsdR subunit of EcoR124I solved by our group (2W00). Motif VI of DEAD box proteins is located at the interface between domain I and II and is important for both ATPase activity and nucleic acid binding. It was suggested that the first and second arginine of Motif VI could bind the gamma phosphate of ATP in the DEAD box proteins and could function like an "arginine finger" that would stabilize the water-Mg^{2+}-beta/gamma phosphate intermediate in the course of ATP hydrolysis.(Caruthers, Johnson *et al.* 2000). Motif Q was discovered in yeast eIF4A and is known to be involved in ATP binding and hydrolysis (Tanner, Cordin *et al.* 2003). Interaction between the bound nucleotide (adenine) and the Q motif has also been reported (Benz, Trachsel *et al.* 1999).

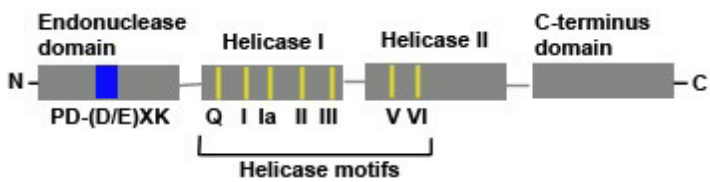


Fig. 1.5: Domain representation of the HsdR subunit of EcoR124I, Grey color denotes domains of protein while yellow represents conserved motifs of helicases the blue rectangle shows X-region which contain the catalytic sites. Figure adapted from McClelland and Szczelkun, 2004.

The N terminal of the motor subunit has regions X responsible for the catalytic activity which belongs to characteristic motif PD-(E/D)xK superfamily (Bujnicki and Rychlewski 2001, Steczkiewicz, Muszewska *et al.* 2012) present in TypeI, TypeII, TypeIII R–M enzymes and RecB

family of nucleases. Several mutagenesis experiments show the relevance of this motif for endonuclease activity of EcoA1 (Janscak and Bickle 2000) and EcoKI (Davies, Kemp *et al.* 1999, Davies, Martin *et al.* 1999). Sequence analysis of the N-terminal region of the HsdR subunit of EcoR124I shows that residues ASP151,GLU 165 and LYS 167 are part of motif II and motif III that together form the catalytic site. They are highly conserved in both EcoAI and EcoKI. Mutations in Motif II and III confirmed the key role in catalysis in RecB family (Yu, Souaya *et al.* 1998, Wang, Chen *et al.* 2000). Sisakova *et al.* did an extended analysis of region X and identified the fourth motif located in a region close to Motif III with the conserved sequence QxxxY. Furthermore, a secondary role in the activity of RecB family nucleases was reported which suggests that the tyrosine residue of the motif could be involved in stabilizing the catalytic domain on the DNA (Sisakova, Stanley *et al.* 2008).

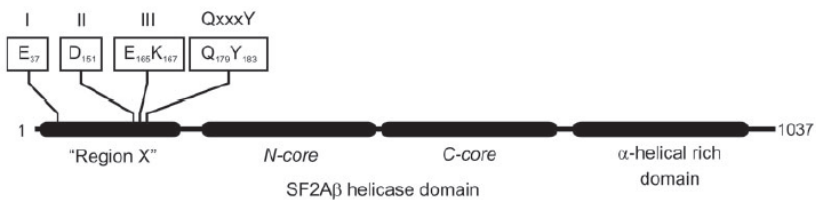


Fig.1.6: schematic diagram of catalytic motif and QxxxY motif in endonuclease domain of HsdR subunit of EcoR124I.

A first structural model of the HsdR subunit was proposed using protein fold recognition method and homology modeling. Obarska *et al.* combined SANS data as restraints to construct a single molecule from models of individual domains. They first defined the domain boundaries by searching conserved domain database and built the individual

domains based on multiple templates. For N-terminal domain, structures of PD-(D/E)xK nucleases (Holliday junction resolvase) were taken as the best modeling templates, while known structures of DEAD-box helicases (UvsW helicase) showed the best alignment with two RecA-like domains. On the other hand, C-terminal domain of HsdR shows no significant similarity to any proteins with known tertiary structures. However secondary structure prediction indicated that this region is rich in helices and also shows the potential to form two to four coiled coils. After building the initial model with the help of the so-called Frankenstein's monster approach, they optimized the structure and tried to fit DNA by overlaying the DNA bound structure and also added the ATP and Mg^{2+} from the known template (PDB ID: 2DB3). Finally they predicted the functional residues responsible for catalytic activities and the residues involved in interacting with the DNA as well. In the end they attempted to determine the overall structure of the HsdR subunit using the information from SANS experiment and revealed considerable conformational heterogeneity within the spatial constraints (Obarska-Kosinska, Taylor *et al.* 2008).

1.4.5 Pentameric complex of EcoR124I

The two HsdR subunits sequentially bind to the Type I methyltransferase to form the complete functional pentameric restriction modification system with the stoichiometry of S_1:M_2:R_2 in EcoKI (Eskin and Linn 1972, Nagaraja, Shepherd *et al.* 1985). Both HsdR subunits have the same binding affinity with methyltransferase. But in case of EcoR124I, the tetramer R_1:M_2:S_1 is predominantly formed and another HsdR binds very weakly (Janscak, Dryden *et al.* 1998). Assembly of the pentameric

complex in EcoR124I from trimeric methyltransferases occurs through a tetramer stable complex with stoichiometry of R_1:M_2:S_1. This tetramer shows ATPase activity but cleavage is only possible when the other HsdR binds. Surface Plasmon resonance study shows that the binding affinity of HsdR subunit with methyltransferase differs by two orders of magnitude (Mernagh, Janscak *et al.* 1998). In case of EcoAI neither of the HsdR subunit binds strongly with the trimeric methyltransferase (MacWilliams and Bickle 1996) while EcoB1 exists in various stoichiometry (Eskin and Linn 1972). Powell and coworkers demonstrate that the complex has a higher binding affinity for any DNA in the absence of AdoMet and ATP irrespective of specificity (Powell, Dryden *et al.* 1993).

Recently, Kennaway and coworkers presented two structures of Type I R–M systems based on electron microscopy together with SAXS/SANS and detailed molecular modeling (Kennaway, Taylor *et al.* 2012). They considered all biochemical constraints, the known crystal structure of the hsdR subunit and position of subunits at low resolution in methyltransferase (Callow, Sukhodub *et al.* 2007, Taylor, Callow *et al.* 2010) and also from previous models (Kennaway, Obarska-Kosinska *et al.* 2009) to construct EcoKI with DNA duplex or DNA mimic anti-restriction protein and EcoR124I with DNA-bound and unbound structure to arrange within the negative-staining microscopy map and suggest the evolutionary link between Type I and Type II R–M systems. They also suggested that there is equilibrium between open and closed conformations of Type I system which depends on the particular system. EcoKI prefers closed irrespective of presence or absence of DNA while

EcoR1241 appears to prefer an open form in the absence of DNA but closed form when DNA bound. They have also proposed the opening and closing cycle for DNA binding (Kennaway, Taylor *et al.* 2012).

1.5 Translocation Activity

The presence characteristic secondary structure motifs, typical for the Superfamily 2 RecA-like helicases and nucleases, allows coupling of ATP hydrolysis and DNA translocation (Davies, Powell *et al.* 1998, Janscak, Sandmeier *et al.* 1999) in the hsdR subunit. However, helicase like activity of DNA strand separation has never been reported for Type I R–M system (Szczelkun 2000).Mutational studies in seven DEAD-box motifs showed that mutants deficient in ATPase activity were also deficient in DNA translocation, which indicates that these two functions are clearly linked together. On the other hand DNA cleavage at remote sites catalyzed by endonuclease domain is completely independent of translocation on DNA (Janscak, Sandmeier *et al.* 1999) and follow a similar mechanism of hydrolysis of the phosphodiester bond of Type II R–M system. However, the HsdR subunit is fully functional only when it binds to the methyltransferase and forms the pentameric complex with a stoichiometry of R_2:M_2:S_1 (Dryden, Cooper *et al.* 1997). Only the pentameric complex shows ATPase activity (Janscak, Dryden *et al.* 1998), DNA translocation (Janscak, Dryden *et al.* 1998, Firman and Szczelkun 2000, Seidel, van Noort *et al.* 2004) and cleavage activity (Janscak, Dryden *et al.* 1998).After binding at the recognition site the whole complex translocates DNA with a speed of several hundred base pair per second (McClelland, Dryden *et al.* 2005) and cleaves DNA remotely (Horiuchi and Zinder 1972) about thousand base pair away

from the binding site (Szczelkun, Janscak *et al.* 1997).The EcoR1241 complex moves along dsDNA (Lohman and Bjornson 1996, Soultanas and Wigley 2001) without any property of strand separation and represents thus a true translocase (Szczelkun, Dillingham *et al.* 1996, Stanley, Seidel *et al.* 2006). Studier and Bandyopadhyay proposed that tracking must be bidirectional and enzyme shows cleavage in-between the two recognition sites roughly half way between the sites (Studier and Bandyopadhyay 1988, Dreier, MacWilliams *et al.* 1996, Szczelkun, Janscak *et al.* 1997).

Experiments using magnetic tweezers (Strick, Allemand *et al.* 1998) and fluorescently labeled triplex forming oligonucleotides attached at a specific distance from the EcoR12I binding site give a detailed picture of how an SF2 helicase DNA motor is loaded onto DNA. Seidel *et al.* demonstrated the initiation termination and re(initiation) steps of DNA translocation and studied the kinetics of HsdR binding. k_{hsdR}, initiation k_{ini}, translocation k_{step}, triplex displacement k_{TFO} and termination of the translocation k_{off} (Firman and Szczelkun 2000)result in a model for the displacement profile. Comparison of these dissociation constants in the presence of ATP suggested that there are different pathways of enzyme disassembly in the absence of ATP, which is most likely due to structural rearrangement of the enzyme during initiation of translocation (Seidel, Bloom *et al.* 2005). A single molecule study reported that translocation is related to the introduction of twist and positive supercoils are generated ahead of the motor subunit. The measured translocated distance of 11 ± 2 bp per supercoils indicates that the two motor subunits

of EcoR124I act independently and translocate along the helical pitch of the DNA (Seidel, van Noort *et al.* 2004).

Several models have been proposed to explain the mechanism of cleavage site selection in Type I R–M systems. According to the cooperative model Type I restriction enzymes bound to its recognition site translocate DNA towards itself simultaneously from both the directions and DNA cleavage occurs at the site where two translocating enzymes collide (Studier and Bandyopadhyay 1988, Dreier, MacWilliams *et al.* 1996) while another model reveals that changes in topology generated by tracking along the groove of DNA helix increase the twist which works as topological barrier that may triggers DNA for cleavage (Szczelkun, Dillingham *et al.* 1996). Janscak and group, based on his experiments adopted the collision model of Studier and Bandopadhyay and reported that a single molecule of Type I restriction enzymes can cleave DNA at any physical barrier that is able to cause a halt or pause in the translocation process. Stalling of translocation of enzymes along DNA can be caused by a Holliday Junction, replication fork, positive super coils or collision of enzymes from different families and is able to induce cleavage. During the process of translocation rotation of DNA produces negative and positive supercoils in expanding and contracting loops respectively (Ostrander, Benedetti *et al.* 1990).However, data shows that the introduction of positive supercoiling into plasmid DNA did not have a significant effect on the rate of DNA cleavage by EcoAI (Janscak, MacWilliams *et al.* 1999).

The whole process of translocation occurs along dsDNA towards the stationary holoenzyme and is coupled with ATP hydrolysis.

Spectrophotometric ATPase assays and analysis of steady-state rates demonstrate that ATP hydrolysis is coupled with bidirectional DNA translocation and each HsdR subunit translocates dsDNA towards the stationary enzyme bound at the recognition sequence. The whole phenomenon hydrolyzes 3.1±0.4 ATP molecules to translocate each base pair (Bianco and Hurley 2005). The same study also reveals that the HsdR subunit transiently disengages from DNA approximately every 515 bp but remains in proximity to the DNA through its binding to the methyltransferase and after engaging again whenever the progress is impeded, the HsdR subunit have an increased opportunity to interact with DNA resulting in DNA cleavage. Results show that 113 molecules of ATP are hydrolyzed per base pair during the process of translocation and cleavage of ssDNA. This value include several cycle of engagement-disengagement prior to, during or immediately after DNA cleavage. For ssDNA, 2.2 ATP molecules hydrolyzed to per base pair translocated per HsdR subunit while 1.6 ATP molecules are hydrolyzed on linear dsDNA for the same (Bianco and Hurley 2005).

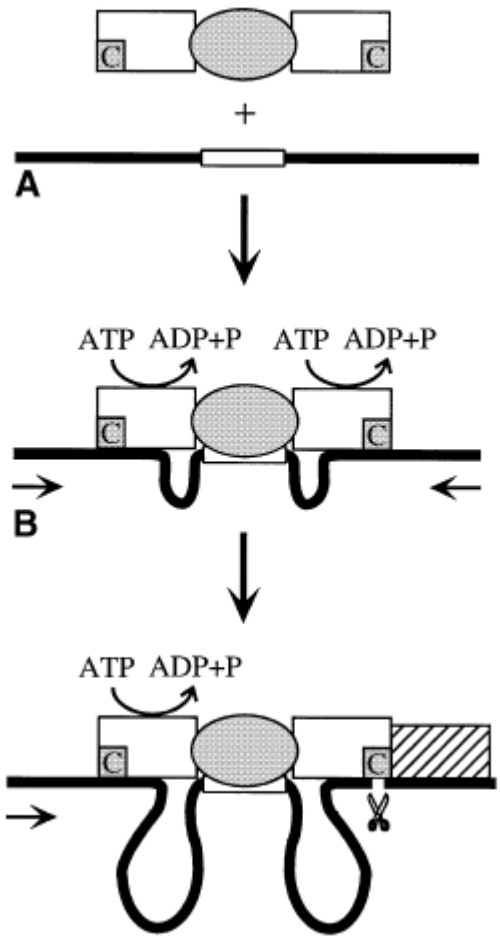


Fig. 1.7: A model for DNA translocation where blockage triggers DNA cleavage. DNA is represented by a dark black line, Methyltransferase by a grey oval and the two rectangles indicate HsdR subunits. Blockage is shown as a hatched box. Figure is adapted from (Janscak *et al.* 1999a).

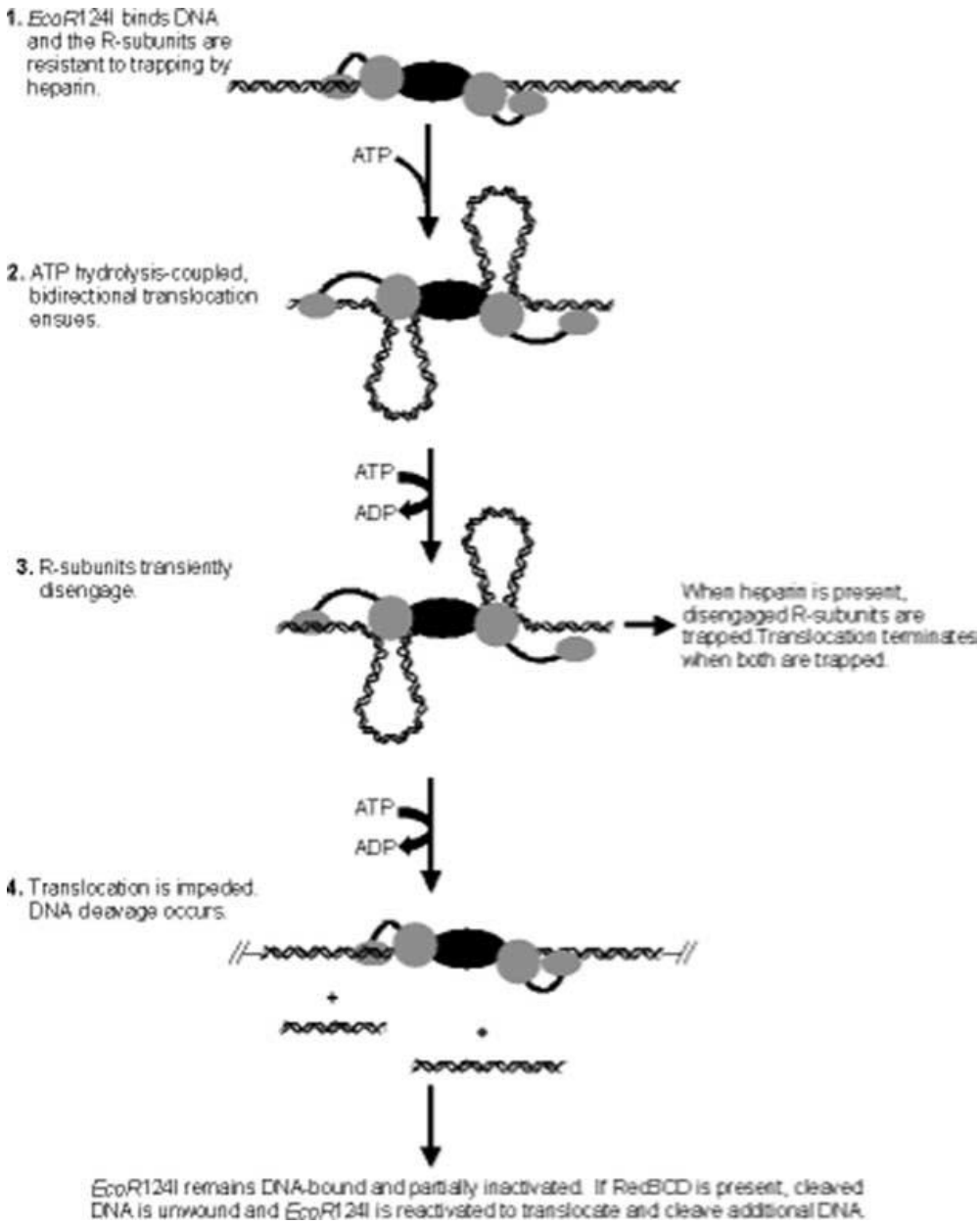


Fig.1.8: Model for DNA translocation in EcoR124I, Figure adapted from (Bianco *et al.* 2005)

1.6 Endonuclease Activity

As previously described, Type I R–M systems cleave unmodified foreign DNA remotely and about thousand base pairs away from the recognition site. During the process of cleavage the enzyme does show no turnover and massive ATP hydrolysis occurs and enzyme does not dissociate from the DNA (Eskin and Linn 1972, Endlich and Linn 1985, Suri and Bickle 1985). The activity of Type I endonuclease is dependent upon the nature of the DNA substrate. Limited or no cleavage activity is found on linear DNA with just a single recognition site (Murray, Batten *et al.* 1973, Rosamond, Endlich *et al.* 1979, Dreier and Bickle 1996).However, if linear DNA contains at least two unmodified recognition sites it gets fully cleaved and cleavage occurs in the region between the two recognition sites if the sites are recognized by the same enzyme (Studier and Bandyopadhyay 1988, Szczelkun, Janscak *et al.* 1997). The cooperative model by Studier and Bandopadhyay explains the reasoning and suggests that DNA cleavage occurs at the site where two convergently translocating enzyme collide. Contrary to linear substrates, circular DNA carrying at least one recognition site readily gets cleaved irrespective of the total number of cleavage sites. Topological restrains during translocation cause positive and negative supercoiling of the circular DNA which stalls the tracking and triggers cleavage (Szczelkun, Dillingham *et al.* 1996). Experimental data suggest that these enzymes can generate both 5′and 3′ overhangs of various lengths. Although EcoKI showed preference for formation of 5′-overhangs of 6-7 nucleotides, and EcoAI for 3′ overhangs of 2-3nt.However, EcoR124I preferentially

formed 5′overhangs of 3-5 nucleotides (Jindrova, Schmid-Nuoffer *et al.* 2005).

It has been observed that EcoR124I cleaves the DNA at both distant *loci* and close to the recognition site which is unique to this particular Type IC enzyme (Szczelkun, Janscak *et al.* 1997). The activity of EcoR12I at *loci* close to its recognition site is reminiscent of the reaction catalyzed by Type III endonuclease. Szczelkun proposed a model to explain this unique cleavage behavior of EcoR124I. According to this model DNA contacts made by HsdS and HsdR subunits produce a relatively small expanding loop. Since tracking might probably require rotation of the DNA and any expanding of loop is only possible by removing the super coiling torque and to overcome the associated energetic barrier most of Type I R–M system nicked the DNA before or just after translocation starts and cleave the second strand later (Bickle and Kruger 1993).Electron microscopy experiments were able to observe this relaxed expanded loop (Endlich and Linn 1985) and translocation proceeds with the help of ATP hydrolysis. As collision occurs either with a second protein or due to a topological constrain, the cleavage site catches the DNA. And at this stage the expanding loop in the stalled complex presents two alternative cleavage sites, one adjacent to the recognition site whose position is constant to the complex and the other coming from anywhere along the length of the DNA molecule and possibility of cleavage at both positions is equally divided. As a result, cleavage in close proximity of the recognition site and other into distant apart occurs.

Another possibility arises when translocation does not start immediately, and then the catalytic site of HsdR subunit introduces a nick in the second strand of DNA before extensive translocation begins. This model suggests the possibility of cleavage of extended loops at close and distant *loci*. However, on a single recognition site of linear DNA the majority of the DNA cleavage reaction occurred close to the recognition site (Szczelkun, Janscak *et al.* 1997).

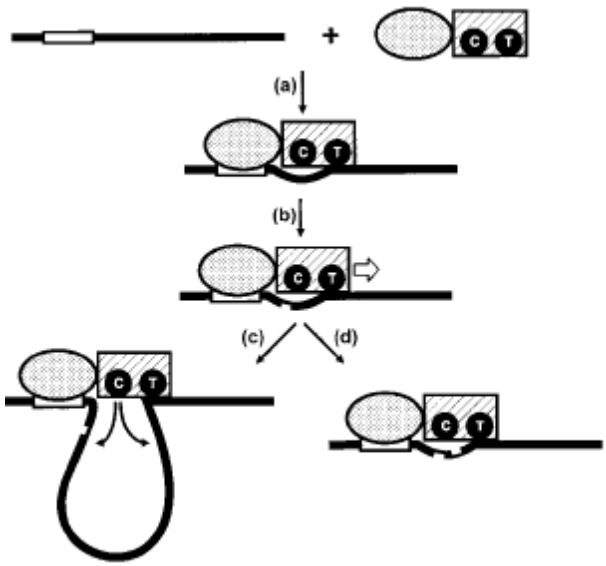


Fig.1.9: Model of Nicking and cleavage of expanding DNA loops by EcoR124I, Figure adapted from (Szczelkun *et al.* 1997).

2. Methods

In this chapter the methods applied to execute the above motivation is described. During this PhD work, detailed atomistic description and molecular interactions were studied as molecular dynamics simulations have shown to be an appropriate tool to compliment the experimental results. Apart from this, several additional tools, methods and algorithm have been used to model, simulate as well as for the analysis of simulation results. The concise description of various algorithm and methods is explained according to their use and importance in the research.

This research work began with the crystal structure of the motor subunit (PDB ID: 2W00) of EcoR1241 which lacks four loops due to its high flexibility. The amino acid residues 142-147, 585-590, and 859-869 were modeled by standard loop modeling procedure in YASARA (Krieger, 2002 #3322). The missing (unresolved) segment for residues 182 to 189 is surprisingly resolved in the Lys220Ala mutant crystal structure (PDB ID: 4BEC) and thus was built into the WT structure by adding the coordinates from the mutant to the above modeled structure.

2.1 Loop Modeling

The most challenging problem in biology is functional characterization of proteins which depends upon the structure. As availability of 3D structure with the help of X-ray and NMR technique is very limited due to complexity of protein structure, comparative modeling is one of the useful techniques that are able to fill the gap between known sequence and known structures to some extent. However functional studies with the help of these predicted structure is still beyond its scope because the functionality is not only determined by shape but dynamics or

physiochemical properties of its solvent exposed and molecular surface also play key role for the particular activity. Usually in the given family substitution, deletion or insertion of residue produces the structural variability and such changes frequently correspond to exposed loop regions that connect elements of secondary structure in the protein fold. Thus loops often determine the functionality specificity of a given protein framework.

Loops are generally too short to provide sufficient information about their local folds. The two main loops modeling procedure have been described: *Ab initio* methods (Fine, Wang *et al.* 1986; Moult and James 1986; Bruccoleri and Karplus 1987) where an energy function is used to judge the quality of the loop. Then this function is minimized using Monte Carlo (MC) (Simons, Bonneau *et al.* 1999) or molecular dynamics (MD) simulation technique (Fiser, Do *et al.* 2000) to achieve the best loop conformation. A knowledge-based technique in which PDB structure for known loops searched with end points that matches the residue between the loops that has to be inserted. Usually much different conformation is obtained for the same set. Possible structures that have to be filtered (Greer 1980)according to geometric criteria or sequence similarity between the template and target loop sequence. (Summers and Karplus 1990, Deane and Blundell 2000).For the short missing segments only fulfilling geometrical parameters are enough to produce reliable conformation. However, several other approaches also that combine these two basic procedures (Chothia, Lesk *et al.* 1986; van Vlijmen and Karplus 1997).

2.2 Computer Simulations

Computer simulation plays a role of interface between theory and experiment. The essence of simulation is the use of the computer to model a physical system which represents the realistic conditions. Molecular dynamics is one of the simulation methods that enables us to follow and understand the structure and dynamics of a given system at the scale where motion of individual atoms can be analyzed in time and space. It was first used by Alder and Wainwright and then by Rahman in the late 1950s and early 1960s to calculate the vibration of atoms in liquids with realistic inter-atomic potential. Due to the revolutionary advancement of computer technology and algorithmic improvements, MD simulations allows us to study different mechanical, thermal, chemical properties of system ranging from fluids, gases, polymers, solids and biomolecules. Since 1970s MD has been widely used to study the structure and dynamics of macromolecules like proteins and nucleic acids (McCammon and Karplus 1977). Biomolecular dynamics occurs on a wide range of time scale and system sizes.

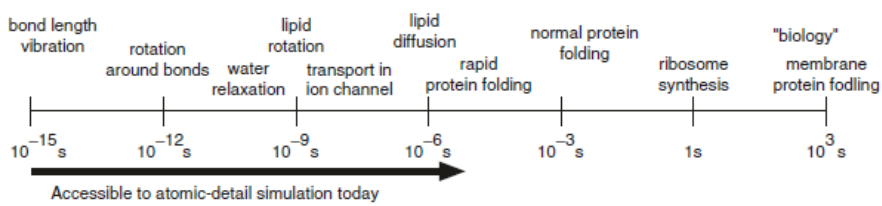


Fig. 2.1: Range of time scales for dynamics in biomolecules.

MD simulations are usually applied for the study of the system of various sizes ranging from few atoms up to several thousand where molecules are treated as classical objects resembling ball and stick model. Atoms correspond to soft balls and elastic springs correspond to covalent bonds. The law of classical mechanics defines the dynamics of the system. It generates the series of successive conformations of a molecular system using Newtonian dynamics. All classical atomistic simulation methods describe physical system as a collection of atoms kept together by inter-atomic forces and the potential energy of the system defined as a function of coordinate of the atoms. The interaction law is specified by the potential $U_{(r1,r2....rN)}$ which represents the potential energy of N interacting atoms as a function of their positions $r_i=(x_i,y_i,z_i)$.

2.2.1 Potential energy terms

The total potential energy of a molecule depends on the arrangement of atoms. In the case of atomistic force field, this is expressed in the terms of mechanical function and can be split into bonding and non-bonding terms.

The bonding terms are bond stretching, bond angle bending, torsion and improper torsion potential. The non-bonding terms are electrostatic and van der Waals potentials. These terms combined to form a potential function which is used for the potential energy calculation of a system.

(i) Bond Stretching Potential

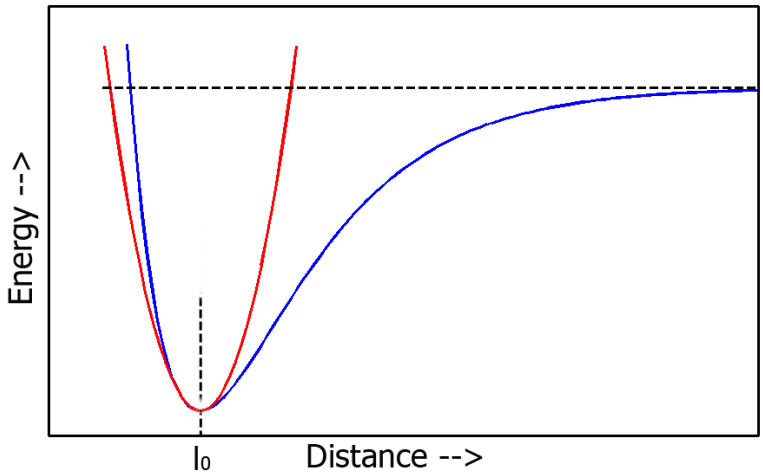


Fig.2.2: Potential energy curve, Morse (Blue) and Hooke's (red) curve.

This is potential energy contribution due to change in the covalent bond length from its equilibrium length. In real case the potential change due to bond length change follows the Morse potential curve. Morse potential models the real stretching but in MM it is generally not used due to complex representation. To make it simpler harmonic potential is used where a bond is assumed as a spring connected between two atoms and it follows simpler Hooke's law representation.

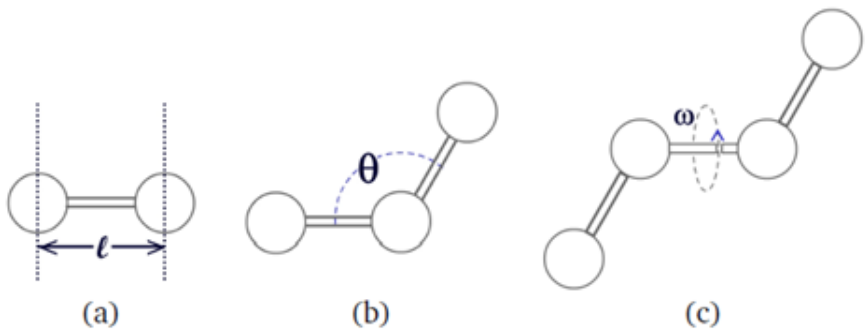
$$V_{bond-length} = K_{bond-length}(l - l_0)^2$$

$K_{bond-length}$ is the Hooke's constant. The value of Hooke's constant is different for different types of atoms. l is the length of bond and l_0 is the

equilibrium bond length. The total bond stretching energy from all bonds is the sum of its contributions from each covalent bond.

(ii) Bond Angle bending Potential

The bond angle between three atoms depends upon the type of hybridization of middle atom, presence of lone pairs, size of atoms and their electronegative properties. For almost all types of atoms optimum bond angle can be determined. This potential energy contribution comes due to the deviation in bond angles from their equilibrium state(s). This potential also follows the Hooke's law like bond length potential functions.



(i) Bond length (ii) Bond angle (iii) Torsion angle

Fig.2.3: Bonded Parameters.

$$V_{bond-angle} = K_{bond-angle}(\theta - \theta_0)^2$$

where $K_{bond\text{-}angle}$ is the Hooke's constant. Θ is the angle and the Θ_0 is the equilibrium angle. The total bending energy from all bond angles is the sum of its contributions from each of the bond angle.

(iii) Torsional Potential

It considers the change in the potential energy due to rotation along bond and manages the interaction of atoms or groups at two ends of the bonds.

As for torsional angle with repulsive atoms/groups at ends, the energy is in the minimum at staggered and maximum at eclipse conformation. The torsional potential can be calculated by the equation:

$$V_{torsional} = V_n[1 + cos(n\omega - \gamma)]$$

where V_n is the barrier height,n is the integer which reprents the periodicity of the minima or maxima from 0 to 2π angle and ω is the torsional angle and gamma γ is the phase modification factor used to assign value of a torsional angle for minimum torsional potential.The biggest movemennt between atoms in molecules are achieved due to torisons and non-bonded interactions.

(iv) Improper Torsion Potential

This potential term is used to maintain the proper geometry of the molecule where any atom or group is attached to a planar sub structure of molecule. It is also called out of plane bending term.

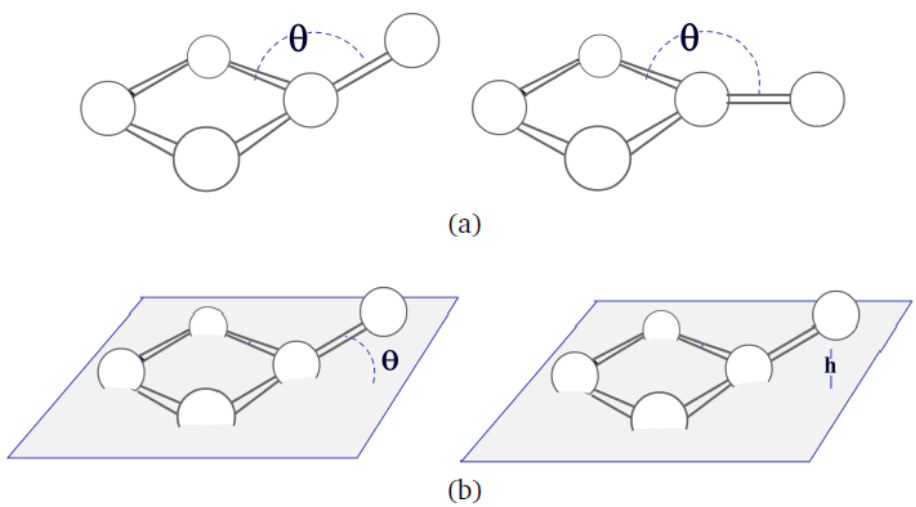


Fig 2.3: Improper Torsion.

A pseudo-torsional term can be used for four atoms where three of them define the plane and the fourth atom makes some angle theta(Θ) with this plane,The potential contribution is then due to change of angle Θ and is given by

$$V_{improper} = k_{improper}(1 - cos(2\theta))$$

where $K_{improper}$ is hardness of the potential and Θ is the angle from the plane created by planner sub structural atoms.

(v) Electrostatic Potential

This is the potential energy contribution due to the electrostatic interaction between charged particles. The total electrostatic potential (ESP) of a system is generally calculated summing up the electrostatic potentials of all pairs of non-bonded charged particles. Generally molecular mechanics treats the atoms as point charges. The charges on atoms are obtained by fitting process to reproduce quantum chemical ESP value (e.g., RESP method is used in AMBER force field.(Wang, Wolf *et al.* 2004). These point charges are used to calculate electrostatic potential using Coulomb's law. For more precise calculation, multipole expansion method can be used, which considers higher order electric moments (i.e., charge, dipole, quadrapole, octapole etc.). The calculation with multipole expansion and is computationally expensive process so only charge-charge interactions (Coulomb's interactions) are considered for general use MM packages.

$$V_{Electrostatic} = \frac{1}{4\pi\epsilon_o\epsilon}\frac{q_i q_j}{r}$$

For i and j particles q_i and q_j are the charges and r is the distance between them, ϵ_0 and ϵ are permittivity of vacuum and the medium respectively.

(vi) van der Waals Potential

van der Waals potential is a widely used expression for the exchange-repulsion and dispersion interactions. This potential is represented by Lennard−Jones(LJ) 12-6 function.

$$V_{vwd} = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}]$$

Where ϵ is the minimum value of potential, σ is the distance between particles where potential is zero and *r* is the actual distance between the particles.

Theoretical calculation and experimental results are used to derive the force-field parameters. The reference bond angle and bond distance are taken from crystallographic data for small molecules while improper dihedral force constant that model the molecular vibrations are obtained from spectroscopic data of small molecules. Dihedral parameters are usually derived from quantum chemical calculations of torsional angle rotational profiles. Partial atomic charges are also obtained from quantum chemical calculations. Parameters for the *LJ* function are usually derived from knowledge of diffusion, viscosity and heats of vaporization of small molecules in the condensed phase and refined by MD simulations. All these parameters have to be derived for all atom types involved in bonded and non-bonded interactions typically present in a bimolecular system.

2.2.2 Force Field

The problem of finding realistic potential that would exactly mimic the true energy surfaces is non–trivial and it leads to many computational simplifications. A typical force field used in the simulations of biosystem is given by:

$$U(\mathbf{r}_1, \cdots, \mathbf{r}_N) = \sum_{\text{bonds}} \frac{a_i}{2}(l_i - l_{i0})^2 + \sum_{\text{angles}} \frac{b_i}{2}(\theta_i - \theta_{i0})^2 + \sum_{\text{torsions}} \frac{c_i}{2}[1 + \cos(n\omega_i - \gamma_i)] + \sum_{\text{atom pairs}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\text{atom pairs}} k\frac{q_i q_j}{r_{ij}}$$

Given the potential, the force acting upon the i^{th} atom is determined by the gradient (vector of first derivative) with respect to atomic displacement:

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1, \cdots, \mathbf{r}_N) = -\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i}\right)$$

Amber force field is generally used in this work and the potential expression for Amber force field is given by:

$$U = U_{stretch} + U_{bend} + U_{tors} + U_{vdw} + U_{electro}$$

More explicitly: $U_{stretch}$ is the bond stretching energy; U_{bend} is the bond bending energy; U_{tors} is the bond torsional energy; U_{vdw} is the van der Waals energy; $U_{electro}$ is electrostatic interaction between point charge.

2.2.3 Molecular Dynamics Algorithm

In MD simulations the time evolution of a set of interacting particles is followed by solving Newton's equation of motion. If $r_i(t)= (x_i(t),y_i(t),z_i(t))$ is the position vector of the i^{th} particle and F_i is the force acting upon the i^{th} particle at time t and mi is the mass of the particle:

$$\mathbf{F}_i = m_i \frac{\mathrm{d}^2 \mathbf{r}_i(t)}{\mathrm{d}t^2}$$

In order to integrate the above second order differential equations the instantaneous forces acting on the particles and their initial positions and velocities need to be specified. Due to many-body nature of the problem the equations of the motions are discretized and solved numerically. The position of each particle in space is defined by $r_i(t)$, whereas the velocities $v_i(t)$ determine by the kinetic energy and thus also temperature of the system.

In MD simulation, the trajectory of the system is the set of the coordinates and velocities $(r_i(t),V_i(t))$ for all the i=1 ,…..N atoms, recoded as a function of time t. In practice, the equation is discretized with a small time-step Δt (typically, in the femtosecond) and the time evolution of the system is obtained by the iterative solution, usually with a "forward" schemes in time, in which coordinates and velocities at time step t+Δt for each atom are obtained from an algebraic combination of coordinates and momenta at time steps t, t-Δt. The most common scheme, which was also used in this work is called leap-frog integrator. The leap-frog algorithm uses positions $r_i(t)$ at the time t and velocities $v_i(t)$ at time $t-0.5\cdot\Delta t$

$$\mathbf{v}_{\mathrm{i}}\left(t+\frac{1}{2}\Delta t\right)=\mathbf{v}_{\mathrm{i}}\left(t-\frac{1}{2}\Delta t\right)+\frac{\Delta t}{m_i}\mathbf{F}(t)$$

$$\mathbf{r}_{\mathrm{i}}(t+\Delta t)=\mathbf{r}_{\mathrm{i}}(t)+\Delta t\mathbf{v}_i\left(t+\frac{1}{2}\Delta t\right)$$

Here $F_i(t)$ is the force vector and can be calculated from potential function as its gradient. The initial coordinates $r_i(0)$ are considered to be at t=0, while initial velocities $\mathbf{v}_i\left(-\frac{1}{2}\Delta t\right)$ are considered to be at time t=-0.5·Δt.

The exact trajectories correspond to the limit of an infinitesimally small integration step. Δt should be small enough so that the gradient of the potential function does not change appreciably within a single time step.

2.2.4 Statistical ensemble and molecular dynamics

In a proper statistical mechanics formulation, each configuration of the atomic system $\Gamma=\Gamma_{(r1,\ldots\ldots rN,p1\ldots..pN)}$ in the (6N-3)-dimensional phase space. (where N is the number of atoms). The time trajectory Γ(t) under the given thermodynamic constraints (constant T,P,V,N) represents the dynamical evolution of the system. According to the Ergodic principle of statistical mechanics the time average of any observable A calculated over the duration τ of the trajectory is equivalent to the ensemble average of A:

$$\langle A\rangle_{ens}=\lim_{\tau\to\infty}\frac{1}{\tau}\int_0^{\tau}A[\Gamma(t)]dt$$

In the limit of infinite length of the trajectory, provided the macroscopic thermodynamics constraints (T,P,V, and E) are the same in the ensemble average and in the time average. The ensemble average of sequence of values of the observable, A_K, depends on the thermodynamics constraints via the partition function Z_{ens}

$$\langle A \rangle_{ens} = \frac{\sum_k Z_{ens}\ (\zeta_k) A_k}{\sum_k Z_{ens}\,(\zeta_k)}$$

Z_{ens} is a function of the parameters ζ_k giving the value of the appropriate thermodynamic potential in the particular configuration k.

The partition function for an isolated system corresponds to the constant (NVE) or *microcanonical ensemble.*

$$Z_{NVE}=\sum_k \delta(H(\Gamma k)\text{-}E)$$

With H=T+V the Hamiltonian (sum of the kinetic and potential energy) of the atomic system, with N atoms in the volume V, calculated in the configuration Γ_k and E the imposed constant energy. The meaning of the Dirac delta function is to select all and only the configurations k whose total energy corresponds to energy E. For a classical system, Newton's equation of motion naturally conserves the energy. Therefore, following a time trajectory Γ(t) in a MD simulation will automatically generate a time order sequence of configurations Γ_k. All with the same total energy $H(\Gamma_k)=E$. In this sense, it is often said that the microcanonical ensemble is the natural ensemble of MD.

Other statistical mechanics ensembles can be useful to perform simulations mimicking particular experimental conditions. For example, the constant (NVT) or canonical ensemble, whose partition function is:

$$Z_{NVT}=\sum_k e^{-\frac{H(\Gamma_k)}{k_b T}}$$

$K_b=1.38 \times 10^{-23}$ J/K b is the Boltzmann's constant and T is the imposed constant temperature. In the canonical ensemble all the possible energy values are allowed, weighted by the Boltzmann's factor, provide the system temperature equal to T. Therefore integration on Newton's equation of motion in a MD simulation would not be a proper way of sampling configurations Γ_k for the NVT ensemble, since these would be constrained to a single energy value.

Similarly the constant NPT or isothermal-isobaric ensemble, is characterized by the partition function:

$$Z_{NPT}=\sum_k e^{-\frac{H(\Gamma)+PV}{k_b T}}$$

With the P imposed constant pressure. In this case the system volume V is free to fluctuate (thus size of the simulation box is also free to fluctuate) during MD simulation in order to adjust to the constant pressure P.

2.2.5 Initial Velocities

This is the initialization step of MD simulation. In this step initial velocities of each atom are generated randomly in the way that Maxwell-Boltzmann's probability distribution for a given temperature is fulfilled:

$$P(v_{\alpha i}) = \sqrt{\frac{m_i}{2\pi k_b T}}\, exp^{-\frac{m_i (v_{\alpha i})^2}{2k_b T}}$$

For each Cartesian component of the velocities $v_{i\alpha}$, where α_i is x_i, y_i or z_i. Before starting the MD simulation the total momentum vector of the system should be zero to prevent translational motion of the center of the mass. These velocities components can be scaled according to the desired initial temperature of the system.

2.2.6 Calculation of the Energy and Force on different particle

This is one of the most time consuming part of the method. Force on each particle can be calculated by differentiation of the molecular mechanics potential function (see molecular mechanics section). Since the most time consuming terms of potential function are pair ways electrostatic and van der Waals interaction, the total numbers of interactions will be $N(N-1)/2$. Where N is the total number of atoms. With more number of atoms, number of interaction terms increases in order of N^2. For a molecule with tens of thousands atoms, it will be very time and memory expensive process for each step. To speed up the simulation, different techniques are available. The most common and efficient technique is to use a cut-off for non-bonded interaction terms as maximum interaction pairs from non-bonded terms. The cutoff term is the radius of the spherically allowed area for non-bonded pair wise interactions. The *LJ* potential decreases very rapidly with distance. At

2.5 σ it becomes 1% of the value at σ. So for *LJ* potential cutoff >=2.5 σ will not exert too much change in the potential energy however when it comes to electrostatic term, no efficient cutoff value can be used. Any reasonable cut-off used for electrostatic potentials shows significant amount of error. In order to compensate this error the long range electrostatic potential calculation, methods like *Ewald* summation and *Particle-Mesh-Ewald* summation have been developed.

2.2.7 Basic work flow of MD simulations

1. Take the initial coordinates of the system.
2. Assign random velocity to each particle of the system using Maxwell-Boltzmann distribution with the adjustment to make total momentum of system zero.
3. Calculate the force on each particle using molecular mechanics force field 4. Using the force vector, calculate next positions and velocities of all particles using Newton's law of motion for next time step($t + \Delta t$).
5. Repeat steps 3 to 4 for desired length of simulation.

2.2.8 Temperature Control

In basic MD simulation the total energy of the system is conserved over the time resulting in a microcanonical ensemble. To perform simulation at constant temperature, we need so called *thermostats*. Thermostats are algorithms which work as an external bath that can exchange heat with the system in order to keep the temperature constant. The basic idea of temperature control is based on velocity rescaling.

Berendsen Thermostat: This is one of the widely used thermostats developed by Berendsen *et al.* in 1984.(van der Spoel and Berendsen 1997)and mostly I used in my work. In this method, the system is coupled with a heat bath of constant reference temperature T_0. The velocities are rescaled at each time step. The rate of change of velocities depends upon the difference of temperature between system and heat bath:

$$\frac{dT_t}{dt} = \frac{1}{\tau_T}(T_0 - T_t)$$

$$\Delta T = \frac{\delta t}{\tau_T}(T_0 - T)$$

Where ΔT is the change in temperature at each time step, Γ_T is the time constant and equal to $(2\gamma)^{-1}$ and is the damping constant. The time constant Γ_T defines the amount of temperature control, as larger the value of Γ_T becomes, weaker will be the coupling. The scaling of velocities of particles is done by scaling factor (λ):

$$\lambda = 1 + \frac{\Delta T}{2\tau_T}(\frac{T_0}{T} - 1)$$

The simple temperature coupling adjustment property makes this thermostat suitable for equilibration steps of simulation. Many other thermostats have been developed like Anderson Thermostat where temperature scaling is done by collision of fictional particles and Langevin Thermostat which uses Langevin equation of Brownian dynamics for temperature scaling.

2.2.9 Periodic Boundary Condition

Computational simulations imitate just a very tiny part of the original system and a very small volume as well. When we see the solvent molecules close to boundary of the box, they would lack the interactions from one or more sides due to absence of molecules in that direction.

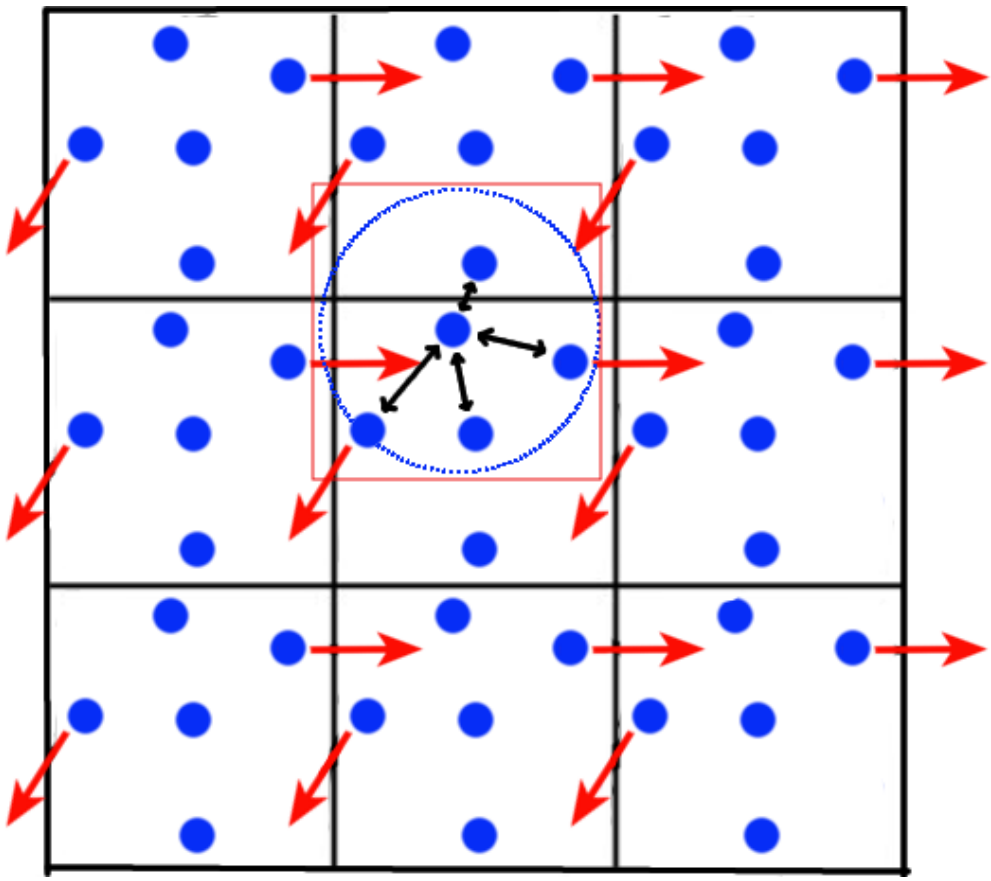


Fig. 2.4: Periodic boundary condition

This would affect the calculation of bulk properties of the system where all molecules feel the interactions from each direction. For the sake of accuracy, boundaries are treated according to the requirement as

"periodic boundary condition" (PBC). PBC is used when simulation is done with the idea to reproduce results for bulk sample. For a system with N particles, the fraction of molecules on the surface of system is proportional to $N-1/3$. To mimic the bulk system, the exact replicas of model system are used and they surrounding the system. If any particle of system moves out from one side of the box then the replica of that particle enters the box from opposite side. In the case of PBC, the particles of one box can interact with the particles of surrounding boxes and this makes appearance of a homogenous system of infinite size. To avoid the interaction of a particle with its own image in neighbor boxes (for minimum image convention), a cutoff of at least half size of the one side length of periodic box should be used.

2.2.10 SHAKE Algorithm

The time step Δt, for solving the Newtonian equation in MD simulation, is governed by the fastest degree of freedom in the system. Generally time step is set to 1 femtosecond which allows modeling of the fastest motion, bond stretching frequency of Hydrogen atoms. Lower frequency modes are responsible for conformation change of the molecules so high frequency modes are less important. Constraint MD is used to increase the step size and to decrease the less important fast motion, thus the time step of the simulation can be increased and thus speeding the calculations. SHAKE is one of most widely used method for constraint MD simulations. In this algorithm, fast degrees of freedom are frozen using holonomic constraints.

2.2.11 Ewald Summation

The use of cutoff and periodic boundary condition bring errors in calculation of long range electrostatic interactions. Ewald Summation is the method for approximation of long-range electrostatic potential. This method splits the electrostatic potential in two parts; the first one for the short range convergent series , which vanishes exponentially with distance (in real space) and second one for long-range convergent series in reciprocal space (Darden T, York D *et al.* 1993)

2.2.12 Solvation

Water is the most common solvent for most biological reactions and therefore it is vital to fully understand water and all of its properties. The complex hydrogen bonding network that water forms can influence protein-protein and protein-substrate interaction and can slow down the protein conformational changes. The network of interactions between individual water molecules affects the energy landscape of proteins. Two major classes of water models are proposed:

In explicit solvent model, calculation the each water molecule is explicitly modeled and each and every interaction within the system required. It is the most computational expensive but provides accurate water model while in implicit solvent model, also known as continuum model, is model where solvent is represented by continuous medium. This representation decreases the computational cost enormously. This method is generally preferred in simulations where the minute details of solvent interactions are not required. Also the absence of friction between solvent-solute molecules in continuum model shows enhanced sampling of phase space.

TIP3P water model is the simplest explicit solvent model developed by Jorgensen *et al.* (Jorgensen WL, Chandrasekhar J 1993). This model represents water molecule by three point charges and rigid arrangement so, no bonded interactions are present. This is one of the simplest models and most common in simple MD simulation. TIP3P model is represented by Coulombic interactions between all pairs of charges and *LJ potential* values between oxygen atoms. TIP3P potential equation for two water molecules is given as:

$$E_{mn} = \left[\sum_i \sum_j \frac{q_i q_j}{r_{ij}}\right] + \frac{A}{r_{OO}^{12}} - \frac{C}{r_{OO}^{6}}$$

where *i* and *j* are point charges in two water molecules *m* and *n* respectively.

Parameters: Charges on hydrogen atom (qH) and oxygen atom (qO) are 0.417and −0.834 respectively. O-H bond length is 0.9572Å, HOH angle is 109.47°, A = 582.0 *kcal*Å*/mol* , and C = 595.0 *kcal*Å*/mol*.

2.2.13 Energy Minimization

In the case of molecular mechanics the multivariate function is represented by the potential energy of the system where the bonded and non-bonded potential terms are the variables. The potential energy terms and their values change with the type of potential function (force-field) used. On the energy landscape view, the potential energy surface (PES) is a multidimensional space. A molecular system of *N* atoms requires 3N-6 internal coordinates or 3N Cartesian coordinates for representation. It is impossible to visualize energy landscape view at once for all

coordinates. Multiple minima, saddle points, barriers, valleys and planes are the components of PES. Minimization process moves the structure to the nearest local energy minima.

It is believed that energy minimized structure representing the optimum structure and good approximation to the native structure of the protein of interest. Several methods have been developed to solve the energy minimization problem. Most minimization algorithms just follow downhill in order to converge to the nearest minimum. These algorithms can be classified into two main classes according to their fundamental ways of operation; first derivative techniques (like steepest descent, conjugate gradient) and second derivative techniques (like Newton-Raphson method). The direction of the first derivative of energy, with respect to coordinates, shows the direction of minima and the steepness is represented by the magnitude of derivative. Second derivative gives information about the curvature of the function. First derivative techniques are slower near the minima but faster when far from the minimum and more robust, so, generally they are followed by second derivative methods which are faster at near minima points. Taylor expansion method can be used to obtain different derivatives of an energy function. Taylor expansion about point x_j can be given as:

$$V(x) = V(x_j) + (x - x_j).\dot{V}(x_j) + (x - x_j)^2\ddot{V}(x_j)/2 + \cdots$$

Where $V(x)$ is the potential energy function of 3N coordinates, $\dot{V}(x_j)$ is $3N \times 1$ sized matrix of partial derivatives (gradient) and (x_j) is $3N \times 3N$ size matrix of partial second derivatives of two coordinates (Hessian)

(i) Steepest Descent method This algorithm is also known as *gradient descent*. Steepest Descent is an example of first derivative minimization algorithm. The minimization process moves in parallel direction of the net force on the system. The direction of gradient is determined by the directions of inter-atomic forces. Direction of the next step can be obtained by $Sj = -gj/|gj|$. If x_j is the current state the new state x_{j+1} will be generated by:

$$x_{j+1} = x_j + \lambda_j S_j$$

Where g is gradient and λ is the step size. The equation is iterated till the energy of new step is not lower than previous one. Each step of iteration takes structure close to the nearest local minimum.

(ii) Conjugate Gradient method This is another first derivative based method. To calculate the direction of next move, this method takes information of direction from the previous step and combines it with the direction of gradient of current step, hence the conjugate gradient method. The use of two gradients increases the efficiency of method by faster convergence than steepest descent method. For a position x_j the direction of movement v_j will be calculated from:

$$v_j = g_j + \gamma_j v_{j-1}$$

$$\gamma_j = \frac{g_j \cdot g_j}{g_{j-1} \cdot g_{j-1}}$$

Direction of movement v_j is used in equation at the place of gradient for the calculation of next step. The process is repeated for defined iteration

number or till the drop in energy is not greater than predefined threshold value.

(iii) Newton-Raphson method- This is the second order minimization technique. This method uses first derivative (gradient) along with the second derivative (Hessian) for the energy minimization purpose. The new state x_{j+1} is calculated by the equation

$$x_{j+1} = x_j - \frac{g_j}{h_j}$$

This method shows faster convergence (lesser number of steps) to local minima than first derivative methods. However, for a large system with hundreds of atoms, the calculation of second derivative of energy with respect to coordinates is computationally extensive task. Another drawback of Newton-Raphson method is that if the system is far from the local minima, minimization may not work properly. Generally this method is used after the first derivative minimization process due to its faster convergence when system is near to any minimum.

2.3 Quantum Mechanics It describes molecules in terms of interactions among nuclei and electrons, and molecular geometry as minimum energy arrangements of nuclei. This method, model the distribution of electrons in molecules explicitly. They are widely used to study the geometries, electronic structure and reactions of small molecules. The system is described by the time independent Schrödinger equations. It is the basis for nearly all computational chemistry methods.

The Schrödinger equation is:

$$\hat{H}\Psi = E \cdot \Psi$$

where $\hat{H}$ is the Hamiltonian operator, Ψ is a wave function, and E the energy. This equation has solution only for particular Ψ called the Eigen function and E known as Eigenvalue. The In practice this equation can be only been solved exactly for one-electron system and different approximation methods are developed for the solution of multiple electron system. In general the wave function Ψ is a function of the electron and nuclear positions and describes a particle as probability wave. This is a probabilistic description of particle (electron) behavior. It can describe the probability of particles (electrons) being in certain locations, but it cannot predict exactly where they are located.

The Hamiltonian operator H for multi-particle system is:

$$\hat{H} = -\sum_{i}^{\text{particles}} \frac{\nabla_i^2}{2m_i} + \sum_{i<j}^{\text{particles}} \sum \frac{q_i q_j}{r_{ij}}$$

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$$

Where ∇ is the Laplacian operator (sum of second derivative) acting on particle i. Particles are both electrons and nuclei. The symbols m_i and q_i are the mass and charge of particle i, and r_{ij} is the distance between particles. The first term gives the kinetic energy and the second term is the energy due to Coulombic attraction or repulsion of the particles.

One way to simplify Schrödinger equation for molecular system is to assume that nuclei move much more slowly than electrons. The Schrodinger equation is then solved for electrons within the fixed nuclei. This is called Born–Oppenheimer approximation, and the Hamiltonian for a molecule with stationary nuclei is:

$$\hat{H} = -\sum_{i}^{\text{electrons}} \frac{\nabla_i^2}{2} - \sum_{i}^{\text{nuclei}} \sum_{j}^{\text{electrons}} \frac{Z_i}{r_{ij}} + \sum_{i<j}^{\text{electrons}} \sum \frac{1}{r_{ij}}$$

Here, the first term is the kinetic energy of the electrons only. The second term is the attraction of electrons to nuclei. The third term is the repulsion between electrons. The repulsion between nuclei is added onto the energy at the end of the calculation. The motion of nuclei can be described by considering this entire formulation to be a potential energy surface on which the nuclei move. Then the wave function Ψ used in the Schrödinger equation is function of electron positions only.

Once a wave function has been d solved any property of the individual molecule can be determined. This is done by taking the expectation value of the operator for that property, denoted with angled bracket< >. For example, the energy is the expectation value of the Hamiltonian operator given by:

$$\langle E \rangle = \int \Psi^* \hat{H} \Psi$$

For an exact solution, this is the same as the energy predicted by the Schrödinger equation.

The simplification used in *ab initio* QM methods is to apply Hartree–Fock (HF) theory, based on approximation that each electron's spatial

distribution is not dependent on the instantaneous motion of the other electrons. In practice individual electrons are confined to functions termed molecular orbitals, each of which is determined by assuming that the electron is moving within an average field of all the other electrons. The total multi electron wave function is written in the form of a single determinant (a so-called Slater determinant). This means that it is anti-symmetric upon interchange of electron coordinates.

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_n(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_n(2) \\ \vdots & \vdots & & \vdots \\ \chi_1(N) & \chi_2(N) & \cdots & \chi_n(N) \end{vmatrix}$$

Here, χ_i is termed a spin orbital and is the product of a spatial function one electron function (molecular orbital), Ψ_i, and a spin function, α or β (also one electron). The set of molecular orbitals leading to the lowest energy are obtained by a process referred to as a "self-consistent field" or SCF procedure. All SCF procedures lead to equations of the form:

$$f(i)\,\chi(\mathbf{x}_i) = \varepsilon\chi(\mathbf{x}_i)$$

Here, the Fock operator f(i) can be written as:

$$f(i) = -\frac{1}{2}\nabla_i^2 + \upsilon^{eff}(i)$$

$\mathbf{x}_i$ are spin and spatial coordinates of the electron i, χ are the spin orbitals and υ_{eff} is the effective potential "seen" by the electron i, which depends on the spin orbitals of the other electrons.

2.3.1 LCAO Approximation

The Hartree-Fock approximation leads to a set of coupled differential equations (the Hartree-Fock equations), each involving the coordinates of a single electron. While they may be solved numerically, it is advantageous to introduce an additional approximation in order to transform the Hartree-Fock equations into a set of algebraic equations.

It is reasonable to expect that the one-electron solutions for many electron system (multi-electron atoms, molecules) will closely resemble the (one-electron) solutions for the hydrogen atom. After all, molecules are made up of atoms, so the molecular orbitals are expressed as linear combinations of a finite set (a basis set) of prescribed functions known as basis function, Ψ.

$$\psi_i = \sum_{\mu} c_{\mu i}\phi_{\mu}$$

c is the (unknown) molecular orbital coefficients, often referred to simply (and incorrectly) as the molecular orbitals. Because the Ψ are usually centered at the nuclear positions (although they do not need to be*), they are referred to as atomic orbitals, and approximation is known as a linear combination of atomic orbitals (LCAO) approximation.

The HF theory ignores electron correlation, the tendency of electrons to avoid each other. The neglect of this effect in the calculation of the total energy has significant implications for chemistry: HF calculations of reactions often give large errors. Many ‘correlated’ *ab initio* methods, including those based on Møller–Plesset perturbation theory (e.g., MP2),

configuration interaction (CI), or coupled cluster (CC) theory, use HF wave functions as a starting point. These methods offer a significant improvement in accuracy over HF calculations, but also have much higher computational cost, which currently makes their application for systems with large number of atoms difficult.

2.3.2 Density Functional Theory

Density functional theory methods can offer accuracy approaching that of the correlated *ab initio* methods, but at substantially lower computational expense. The basis of DFT is that the ground-state energy of a molecule can be calculated just from the knowledge of electron density distribution. The density is a function of only three variables and is thereby much simpler than the *ab initio* wave function, a function of 3N variables, where N is the number of electrons. However, the exact form of the functional relating the density to the energy is not known. Numerous approximate functions have been developed based on a mixture of trial and error and known limiting features of the exact functional, but there is (as yet) no systematic way to improve them. One popular density functional is B3LYP, termed a 'hybrid' functional, in which a degree of HF exact exchange is mixed with contributions from other functional, including the Becke88 exchange functional and the Lee–Yang–Parr correlation functional.

3.Results & Discussion

3.1. Interdomain communication in the endonuclease/motor subunit of the type I restriction – modification enzyme EcoR124I

3.1.1 Introduction

As already mentioned in the motivation part,the endonuclease domain surprisingly makesa contact with ATP in the crystal structure of motor subunit/HsdR subunit (2W00) of EcoR124I.In particular, the amino group of Lys at the position 220 makes a contact with the N3 atom of the adenine ring of ATP which is around 20-Å away from the catalytic site of the endonuclease domain. Such a type of contact is atypical in other RecA helicases or translocases and so far unique for EcoR124I.

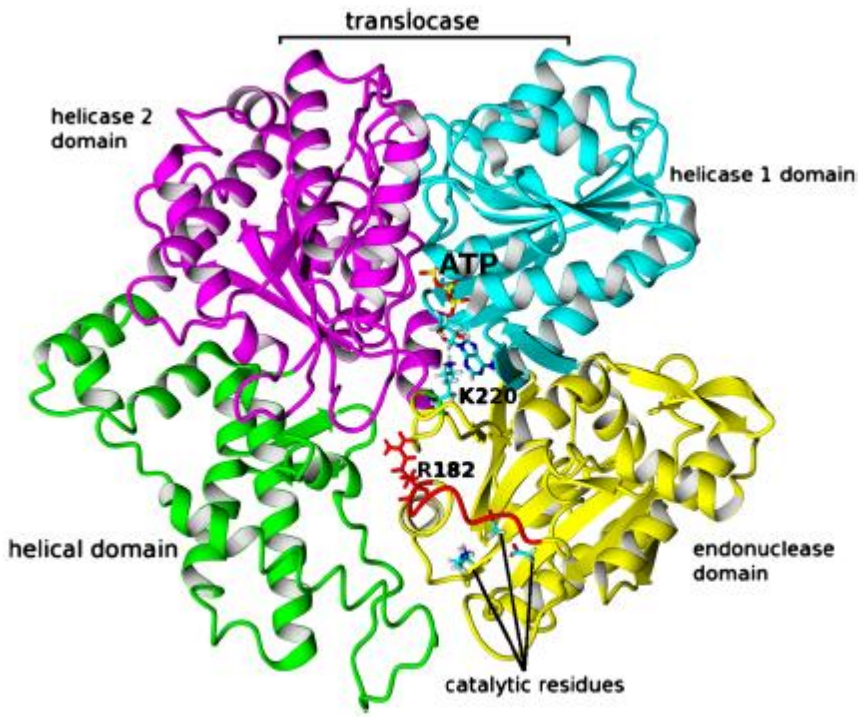


Fig.3.1.1: Crystal structure of the motor/HsdR subunit.Selected side chains relevant for this work is also shown and QxxxY motif is shown in red color.

This contact suggests a potential means for coupling of helicase and endonuclease activity when enzymes becomes viable or functional. Although in the WT crystal structure segment 182-189 is absent/not resolved but in mutant Lys220Ala of HsdR of Ecor124I this segemnt is resolved. Merging computationally the segment 182-189 coordinates from this mutant with the structure of WT, it was found that Arg182 is located right above the $QxxxY_{183}$ motif. As Siskova earlier reported a functional role of Tyr183,initial analysis of the simulations of WT focused on the behavior of Tyr183. Interestingly, we found significant quite correlated conformational change in Arg182 with the Lys220-ATP contact. Since the crystal structure represents only a single snapshot and in action the motor subunit undergoes large conformational changes, with the help of MD simulations we fill out the gap by exploring possible conformation changes. This allowed to describe the role of Arg182as a switch between cleavage-competent and translocation-competent conformations on a molecular level. Detailsareavailable in **Paper I** (Sinha, Shamayeva *et al.* 2014)

3.1.2 Results

Before analyzing global conformational changes inside the protein we initially wanted to confirm the contribution of Lys220 to ATP binding, excluding that the close distance would be by chance only or even a crystallization artifact.To calculate the contribution from the side chain of Lys220, QM/MM studies were performed and comparedto contributions of other residues in the ATP binding pocket. In QM/MM calculations we take ATP and Mg^{2+} as QM part and the rest of the

protein as MM region.To calculate the contribution of each residue in the ATP binding pocket each contacting residue was replaced by a glycine, as glycine doesn't have a side chain.

The contribution of each side chain was calculated as the difference between the summed electrostatic, van der Waals, and solvent contributions when the side chain is present and when it is replaced by a hydrogen atom. The total binding energy of ATP is −118.5 kcal/molwith nine residues contributing significantly to the interaction. In each of these residues the side chains were replaced in turn by a hydrogen atom to determine its individual contribution. The majority of the ATP binding energy is contributed by contacts withthe Mg^{2+} ion and phosphate from six residues (Arg691, Asp408, Glu409, Lys313, Arg688, and Asp664), which each contribute total energies ranging up to −91 kcal/mol. One residue, Arg691, also contacts the ribose ring with the energy contribution ofapproximately−40 kcal/mol. Two residues (Gln276 and Lys220) contact the adenine base, with contributions of −23.9 kcal/mol and −16.2 kcal/mol, respectively. Thus, the energy contribution of Lys220 toATP binding is of a magnitude comparable to that of the onlyother contact to the adenine ring.This indicates that the Lys220–ATP contact is likely to be functionally relevant rather than afortuitous contact or an artifact of crystallization.

Table 3.1.1: Energy contribution (kcal/mol) to the ATP binding of aminoacid residue in the binding pocket.

	ΔE_{elec}(QM/MM)	ΔE_{vdw}(MM)	ΔE_{solv}(MM)	$\Delta E_{binding}$
WT	−480.8	−51.9	414.2	−118.5
Lys220[a]	−91.8	−1.3	76.9	−16.2
Gln276[a]	−17.8	−1.4	−4.7	−23.9
Lys313[b]	−160.7	−0.5	70.0	−91.2
Thr314[b]	−33.0	1.0	−19.8	−51.8
Asp408[b]	52.9	−0.5	−44.7	−7.7
Glu409[b]	83.1	−0.5	−117.8	−35.2
Asp664[b]	65.7	−0.19	−57.4	8.1
Arg688[b]	−129.5	0.8	100.5	−28.2
Arg691[c]	−149.2	−3.1	112.4	−39.9

[a]Residue in contact with the adenine ring
[b]Residue in contact with phosphate/Mg^{2+}
[c]Residue in contact with the ribose

During MD simulations of WT, residue Tyr183 always maintains the outward orientation and never was the hydroxyl of this residue observed to get closer than 8Åto the nearby segments of the endonuclease domain. This leaves open a straightfoward explanation of the pronounced effect of the Tyr183Ala mutation *in vivo* and *in vitro*.However, during the inspection of the behaviour of Tyr183 in these trajectories it was observed that the behvior of the adjacent residue is strongly coupled to the behavior of the loop bearing Lys220.This observation led to the detailed analysis of Arg182 in these simulations that is presented below; the predictions then led to wet experiments allowing analyses of the behavior of an Arg182Ala mutant HsdR *in vivo* and *in vitro*; and to consequent simulations with this Arg182Ala *in silico* mutant.

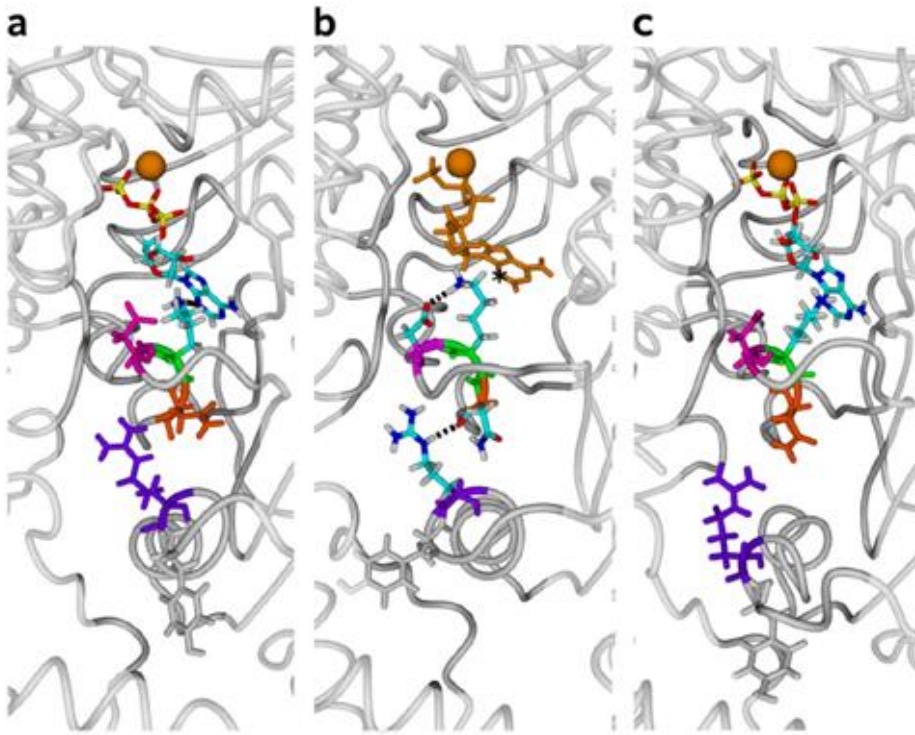


Fig.3.1.2: Local interactions in the ATP-binding site. The center of the planar array of the four HsdR domains is shown in a zoomed-in view with Mg^{2+}-ATP in the top center. This viewpoint was chosen to visualize as many of the relevant interactions as possible, and similarly in all three panels.

The polypeptide chain in **Fig.3.1.2** is shown as a gray tube, and residues involved in hydrogen-bonding interactions discussed in the text are shown as sticks. Because some hydrogen bonds (black dashed lines) are obscured regardless of viewpoint, residues of interest are identified by unique solid colors when they are not involved in hydrogen bonding, or atomic colors with cyan carbons when they are involved in hydrogen bonds, except for their Cα atoms, which retain the unique color in order to permit the identification of each residue. Gold Mg^{2+}-ATP, orange Asp219, green Lys220, magenta Asn221, purple Arg182, gray Tyr183.In

the figure (a) shows the initial model structure used for simulations of WT HsdR, prepared from the X-ray crystal structure of HsdR by adding four short loop structures as described in the text. The hydrogen bond indicated between the Lys220 Nζ and adenine N3 atoms is inferred from the distance and angle in the crystal structure. (b) represents a typical snapshot from the equilibrated part of the simulation of WT HsdR. AdeN3 is marked with an asterisk; and (c) represents a typical snapshot from the equilibrated part of the simulation of restrained WT HsdR.

Analysis of Lys220 The initial hydrogen-bonding interaction between the Lys220 ε-amino group (Nζ) and adenine ring atom N3 (AdeN3) observed in the crystal structure is rapidly lost and not regained in all production runs of the WT HsdR structure analyzed in this work. Other ATP interactions with comparable energies are maintained; in fact, no other ATP contact is lost during these simulations. QM/MM calculations indicate that this interaction (interatomic distance Lys220Nζ to AdeN3 is 3.09 Å in the published crystal structure, contributing approx.−16 kcal/mol) is of similar magnitude to, e.g., Gln276 (approx.−24 kcal/mol). The mean±s.d. distance between Lys220Nζ and AdeN3 during the equilibrated last 50 nsof the WT simulation is ∼7±∼1 Å (**Fig. 3.1.4b**). Instead ofinteracting with AdeN3, the Lys220 side chain turns towardthe adjacent Asp219 and forms a hydrogen bond with itsside-chain carboxylate Oδ1 atom that persists for theentire simulation. The motion of the 220s loop is in thedirection of the 180s loop, and brings the backbonecarbonyl O atom of Asn221 within hydrogen-bondingdistance of the Arg182 guanidino Nε atom, from aninitial distance of 3.24 Å to a mean equilibrated distanceof ∼2.9±0.3 Å. A hydrogen bond is formed

betweenthese two atoms that persists for the entire simulation.The distance between Lys220Cα and Arg182Cα duringthe equilibrated part of WT simulations is ~11±~0.5 Å,with the respective side chains getting never closer to eachother than ~9.5 Å.

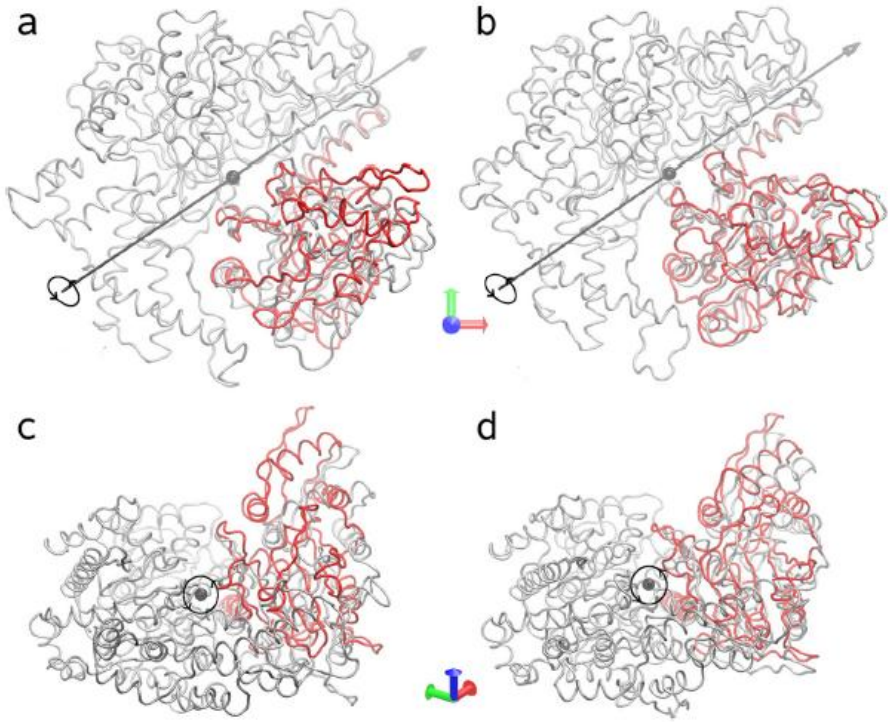


Fig. 3.1.3: Endonuclease domain rotation. The planar array of four HsdR domains is shown in tube representation. The coordinate system used to define the rotation is indicated at the center of the panels: red x-axis, green y-axis, blue z-axis. The vector describing the rotation of the endonuclease domain is shaded in black to gray from near to far, and the direction of rotation is indicated by the circle arrows at the origin of the vector. For the sake of orientation in Figs. a and b, the center of the vector shown by the gray sphere is approximately the position of the bound Mg^{2+} ion. (a) The average conformation in the equilibrated part of the restrained simulation ofWT HsdR is shown in gray. The conformation in the equilibrated part of the WT HsdR simulation is overlaid and, except for the endonuclease domain (red), it is not shown because it is identical. (b) Extreme positions along the first eigenvector calculated by principal-components analysis for the second half of the

simulation of Arg182Ala HsdR. The extreme shown in gray corresponds to the conformer at 56 ns. When it is overlaid on the other extreme corresponding to the conformer at 39 ns (not shown), it is identical, except in the endonuclease domain (red). (c) As in panel a with a view along the screw axis from the near end. (d) As in panel b with a view along the screw axis from the near end.

Loss of the Lys220–ATP contact involves not only the movement of the Lys220 side chain and the 220s loop but also rotational movement of the entire endonuclease domain relative to the other three domains (**Fig. 3.1.3**a, c). Principalcomponent analysis shows that the domain rotates by ∼18° about a screw axis defined by the vector (0.75, 0.5, −0.43) through the center of mass of the endonuclease domain in a coordinate system in which the domain is positioned as in **Fig.3.1.3**, with the z-axis orthogonal to the plane of the page. This screw axis applies only to the endonuclease domain and not to the other three domains, which show no rotational motion and stay in place during the simulations, as is evident in **Fig. 3.1.3** (a and b). This rotation is in the same direction,although it is much smaller in magnitude than the rotation of the endonuclease domain as observed in the X-ray crystal structure of a fragment of a putative HsdR protein from *Vibrio vulnificus* (PDB ID: 3H1T).Relative to the crystal structure of EcoR124I HsdR subunit, the *Vibrio* endonuclease domain is rotated by ∼180°; the screw axis would be described by the same vector, but with the axis shifted toward the center of mass of the endonuclease domain, i.e., downward to the lower right in **Fig. 3.1.3a** Although the *Vibrio* HsdR polypeptide chain was incomplete, the structure resolved both helicase domains and the entire endonuclease domain, and thus suggests that a range of motion is possible for the latter domain that is much greater than that observed in the present simulations.

Analysis of Arg182 The conformation of the 180s loop borrowed from the Lys220Ala mutant structure, when modeled into the WT structure, places the 180s loop in proximity of the 220s loop, where Lys220 is engaged with ATP. The loss of the Lys220–ATP contact during simulations starting from this structure could reflect that this conformation of the two loops might be incompatible. The persistence of the Arg182–Asn221 interaction during the entire WT simulation upon the loss of the Lys220–AdeN3 contact further suggests that these two interactions may reflect alternative states of the system. Thus, to enable observation of the behavior of the 180s and 220s loops while ATP is engaged by the endonuclease domain (the state of the enzyme detected crystallographically), additional simulations were prepared and run in which the Lys220–AdeN3 interaction was maintained artificially via a restraining force. Due to technical difficulty in applying an intermolecular distance restraint in Gromacs, the restraint was accomplished by identifying the backbone nitrogen atom of Arg273 as a locus where an applied distance restraint to Lys220Nζ would keep the Lys220 ε-amino group within hydrogen-bonding distance of AdeN3. The force constant of the distance restraint was incremented until the Lys220Nζ–AdeN3 interaction was maintained at a distance of ∼3.5±∼0.5 Å during the entire simulations, which are hereafter referred to as restrained WT simulations. The final force constant used was 5,000 (kJ/mol nm^2). A representative snapshot from the restrained WTsimulation is shown in **Fig.3.1.3**c. The position of the endonuclease domain corresponds to the crystal structure and no global rotational movement is observed, i.e., rotation about the screw axis in **Fig. 3**b and d is 0°. In these simulations, the distance from Arg182Cα to Lys220Cα is

~15± 0.5 Å, with the respective side chains coming no closer to each other than ~13.5 Å.

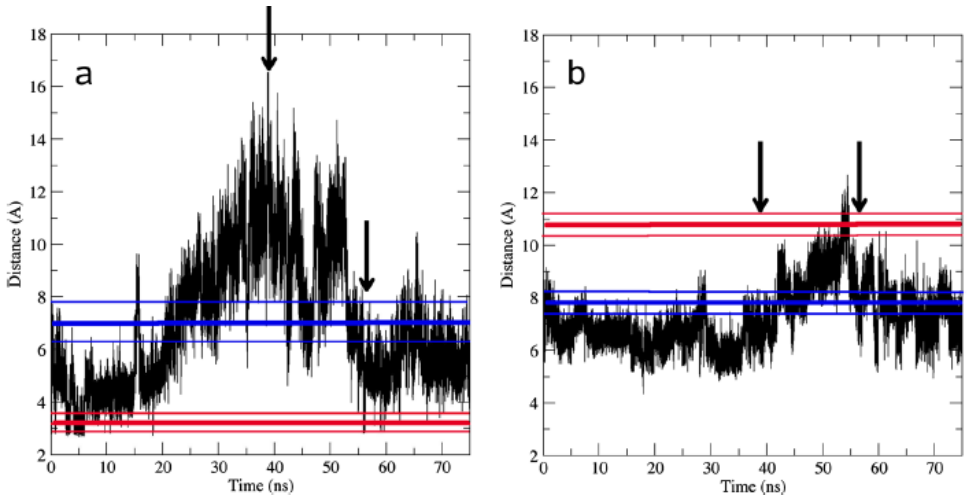


Fig. 3.1.4:a–b, Correlated and uncorrelated loop motions. Interatomic distances are plotted as a function of simulation time during the Arg182Ala mutant simulation (black traces). Horizontal heavy lines are the mean distances over the equilibrated part of the WT simulation (blue lines) or the restrained WT simulation (red lines). Lighter horizontal lines mark one standard deviation from the mean distance. Conformers corresponding to the extremes along the first eigenvector in the principal-components analysis are indicated by arrows: 39 ns rotated conformer, 56 ns unrotated conformer. a Lys220Nζ–AdeN3. b Arg182Cα–Asn221Cα

Fig 3.1.4 compares relevant distances that illustrate themovements of the 180s and 220s loops in the final equilibratedstructures of theWT (blue lines) and restrainedWT (red lines)simulations. Panel a shows the mean±s.d. distance betweenLys220Nζ and AdeN3 in the two simulations, and panel bshows the mean±s.d. distance between Asn221Cα andArg182Cα. When the Lys220–ATP contact is engaged, as inthe restrainedWTsimulation (red line in panel a), the 180s and220s loops are far apart (red line in panel b), with a welldefinedand narrowly

distributed distance between Asn221Cαand Arg182Cα (~10.8±~0.4 Å). This distance is too large topermit atomic contacts between the loops, and indeed nohydrogen bonds or salt bridges between them are detected inthe restrained WT simulation. On the other hand, when the Lys220–ATP contact is lost and Lys220 forms a persistentinteraction with Asp219, as in theWTsimulation (blue line inpanel a), the 180s and 220s loops are in close contact with anarrowly distributed distance from Asn221Cα to Arg182Cαof ~7.9 ± ~0.5 Å. The finding that loss of the Lys220–ATPcontact is coupled to domain movement, together with thefinding that the behavior of the 180s loop is coordinated viaArg182 with the behavior of Lys220, suggests a role for Arg182 in communicating the ATP-ligation state through theendonuclease/motor subunit.

Ala182: To test this hypothesis, a mutation of Arg182 to Alawas introduced both *in silico* and by *in vitro* mutagenesis. Theinitial conformation of the virtual mutant is identical to theinitial WT conformation except for the missing Arg182 sidechain, which was replaced by in silico mutation with a methylgroup. Simulations using the Arg182Ala virtual mutant structurewere prepared and run for at least 75 ns, with equilibrationduring the final ~20 ns as judged from the RMSF values are similar tothose reported above for WT and restrained WTsimulations.The Lys220Nζ–AdeN3 and Ala182Cα–Asn221Cα distancesin the Arg182Ala simulation are superimposed ontothe WT and restrained WT simulation distances in Fig. **3.1.4** (black traces). As in WT, the contact of Lys220 with ATP islost early in the Arg182Ala simulation (panel a). This resultindicates that the 220s loop moves away

fromATP even whenthere is no long charged side chain at residue 182. Themovement is toward the 180s loop initially, and for the first∼30 ns the simulation approximately follows the course of theWT simulation. However, after ∼30 ns the distance between Lys220Nζ and AdeN3 increases further, reaching a maximumof ∼17.4 Å at ∼40 ns. The fact that the Lys–ATP distancecontinues to increase beyond the distance observed in theWTsimulation suggests that the 220s loop enjoys more degrees offreedom than in the WT structure, perhaps because of theabsence of the Arg182 side chain that in the WT simulationinteracts persistently with Asn221.

This picture of the role of Arg182 is reinforced by analysisof the Ala182Cα–sn221Cα distance (**Fig. 3.1.4b**), which showsthat the two loops explore a range of distances but with nocorrelation in time between their movements. Thus, in theArg182Ala mutant, the interloop distance is insensitive tothe contact between Lys220 and ATP. As in the WT simulation,the Arg182Ala simulation also develops a pronouncedrotation of the endonuclease domain (maximum indicated byarrow at 39 ns) about the same screw axis as in the WTsimulation, but the rotation is only ∼11° (**Fig. 3**b and d). Therotation is not found in the equilibrated part of the m utantsimulation, where the rotation is zero, but occurs only in thefirst 40 ns, when the system is not yet equilibrated. Themaximumrotation of the endonuclease domain coincides withthe maximum in the distance between Lys220Nζ andAdeN3 at the ∼40-ns time point. Thus, the degree of rotationis also insensitive to the contact between Lys220 and ATP.

In the second half of the Arg182Ala simulation, Lys220slowly returns to a position close to the initial one, and after∼55 ns it approaches ATP again. During this time period,rotation occurs in the opposite direction, toward the nonrotatedposition of the endonuclease domain observed in thecrystal (arrow at 56 ns). During the short (20 ns) equilibratedpart of the simulation that could be observed, the Lys220Nζ–AdeN3 hydrogen bond is not fully re-established and hydrogen bond distances are seen in only a few snapshots.

These results indicate that the rotated conformation with the maximum Lys220–ATP distance does not persist in the Arg182Ala mutant simulation as it does in the WTsimulation. Thus, the mutant is apparently not trapped in the alternative rotated conformation that traps theWTstructure. These results for the Arg182Ala virtual mutant indicate that Arg182 plays a strong role in the switch-like behavior observed in WT simulations, and that this role involves communicating the ATP ligation state through the endonuclease/motor subunit by controlling a switch between mutually exclusive alternative structures. The rotational transition in the first half of the Arg182Ala simulation, and its reversal as the system equilibrates, raise the question of whether the presence of the Lys220–ATP contact is sufficient to prevent the endonuclease domain from rotating, or if the contact between the 180s and 220s loops plays a role. The first 40 ns of the WT simulation suggest that loss of the Lys220–ATP contact initiates domain rotation. Mutation of Arg182 does not alter this behavior, and after 40 ns the endonuclease domain is rotated in both simulations. However, in the mutant, the rotated state does not equilibrate as in WT, but after 50 ns the domain rotates in the

opposite direction in concert with the movement of Lys220 toward ATP. This result suggests that a factor stabilizing the rotated conformation is present in WT HsdR but absent in the mutant, and that this factor involves Arg182. Comparison of panels a and b shows that in the Arg182Ala virtual mutant the 180s to 220s interloop distance changes independently of the Lys220–ATP distance and the endonuclease domain rotation, whereas in the WT simulation the interloop distance is dependent on the Lys220–ATP distance. The presence of the Lys220–ATP contact therefore appears to be insufficient by itself to prevent endonuclease domain rotation unless a restraining force is applied to maintain it or the contact made by Arg182 with the 220s loop in the rotated conformation is absent.

Conclusion: The above results reveal the importance of R182 in inter-domain communication as it plays a key role in communicating ATP-ligation to the endonuclease catalytic sites. Conformational changes in the 180s loop and 220s loop are correlated and coupled which a rotation of the endonuclease domain as a whole.

(To view experimental results and discussion please see Paper I (Sinha, Shamayeva *et al.* 2014)).

3.2 The helical domain of the motor subunit of EcoR124I participates in ATPase activity and dsDNA translocation

3.2.1 Introduction

The process of translocation is driven by two RecA-like helicase domains and consumption of one ATP per base pair. It has been observed that large conformational changes occur in the helicase II domain and during the translocation cycle the enzyme passes through several stages like intake of ATP, ATP hydrolysis release of ADP and resetting the structure for intake of ATP again. Rad54 dsDNA translocase (SF2 superfamily) from *Sulfolobus solfataricus* was proposed to translocate DNA in an inchworm fashion (Dürr *et al.*, 2005). Thereby, binding to DNA in the absence of ATP leads to a rotation of helicase domain 2 with respect to helicase domain 1 by 180° (Dürr *et al.*, 2005; Lewis *et al.*, 2008). Although the binding of the protein to the DNA is strictly speaking not part of the regular mechano-chemical cycle and it remains unclear if this large conformational change is physiologically relevant and occurring *in vivo*, it demonstrates the ability of the second helicase domain to undergo large rotational movements (Lewis *et al.*, 2008).

The reported crystal structure of HsdR of EcoR124I represents one of the ATP bound stages in the translocation cycle. In this crystal structure the helicase2 domain has an inter-domain contact with the c-terminal helical domain at the inter-domain interface. The presence of such inter-domain interactions at the interface raises the question whether these are actively participating in the translocation cycle of the translocase domains, as they could potentially stabilize a particular translocation stage.

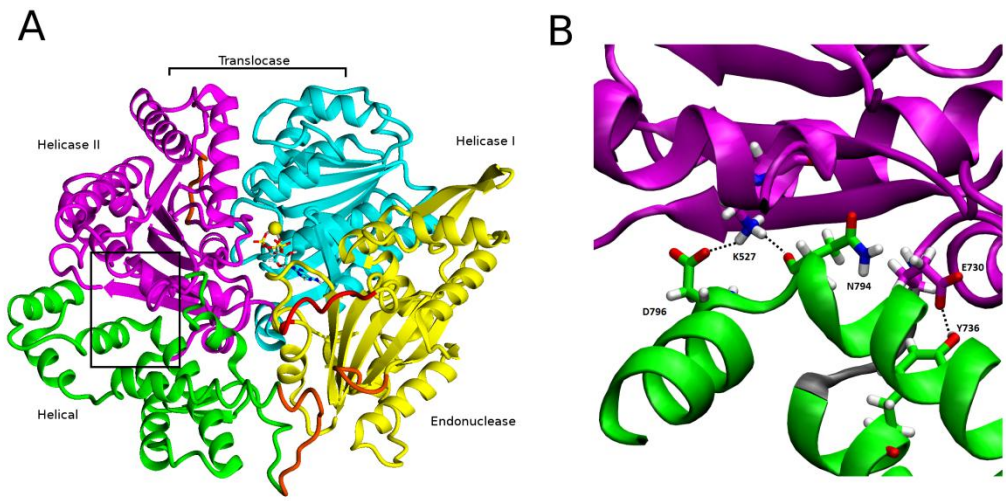


Fig.3.2.1. A. Structure of HsdR consisting of four domains which are color coded here, Yellow is the endonuclease domain while Helicase 1 and Helicase 2 domains are depicted cyan and magenta respectively. The helical domain is shown in green. ATP is represented in elemental color as a skeletal model and magnesium is indicated by a blue sphere. The modeled parts of the structure (not resolved in the crystal structure) are in red color. **B.** Residues involved in hydrogen bonding at the interface between the helical domain and helicase domain 2 are shown as a skeletal model: the amino group of LYS527 of Helicase 2 domain contacts the carboxyl oxygen of ASP796 and the backbone oxygen of ASN794 on the helical domain while the hydroxy group of TYR736 contacts the carboxyl group of GLU730. All interactions are indicated with black dotted lines.

A closer look to the crystal structure of the HsdR subunit of EcoR124I (Lapkouski *et al.*, 2009) reveals one contact centrally located between the helicase 2 domain and the helical domain: a bifurcated contact of the K527 amino group in which the K527 heavy atom Nζ and the D796 carboxyl Oδ atom are in a 2.9 Å distance, while K527 Nζ is 2.7 Å from the N794 backbone oxygen. A second contact is observed in the helix-turn-helix domain connecting region (linker region). This linker region between helicase 2 and helical domains is formed by α-helices 32 and

33, where Y736 contacts E730 with the phenol O of Y736 and atom Oε2 of E730 being in a 2.6-Å distance, potentially stabilizing the bend/hairpin in the linker.

In this part of my thesis, the relevance of the above-mentioned interactions for the translocation activity of EcoR124I is evaluated employing MD and predictions are confronted with analysis of mutant enzymes by *in vivo* and *in vitro* experiments. The reported mutations are the first reported mutations in the helical domain that result in a functional consequence other than a loss of the ability to assemble the full complex.

3.2.2 Results

Global conformational changes

Global motions during the simulations were analyzed by principal components analysis (PCA). Equilibrated trajectories from the last 20 ns of each individual WT and mutant simulation were extracted and finally joined into one long trajectory. Running the PCA over the combined trajectory allowed the identification of the first eigenvector in the system (WT or mutant), and to which degree this largest conformational change and identical motion occurs in the individual proteins, WT or mutant. The motions along the first eigenvector were identified and describe by overlaying three-dimensional structure after 100 ns of the molecular dynamics simulations with the initial structures (**Fig. 3.2.2**) which reveal the global collective motion described by the first eigenvector to be a partial opening of the cleft between helicase 1 and helicase 2 domains in mutants K527A, Y736, and K527A_Y736A. The mutant D796A

maintains the same conformation as observed in the WT, and the structures before and after the simulation. This result suggests that in the D796A mutant, K527 is able to keep a persistent contact with the helical domain even though D796 is not available as a partner anymore.

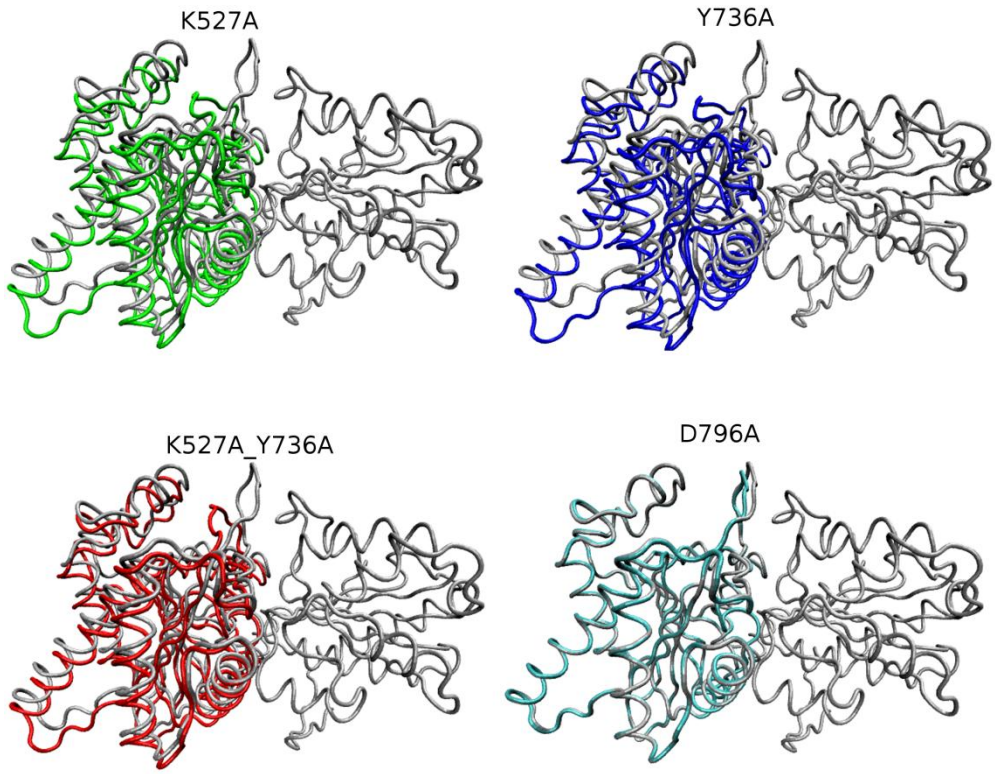


Fig. 3.2.2: Overlay of WT and mutant structures. Both domains of the translocase (Helicase 1 and Helicase 2 domains) are shown in tube representation. The WT conformation of both domains is shown in gray. Mutant conformations are fitted to helicase domain 1, which is thus identical in all cases and therefore shown only once for better visualization, while the position of the helicase 2 domains is depicted in respective colors.

To quantify these structural changes, projection along the first eigenvector analyzed, these projections are described by the coordinate along the first Eigenvector **(Fig.3.2.3).** This coordinate represents the

collective motion of all Cα atoms. The WT HsdR does not show any major change and its projections value oscillates around zero, as expected for WT because the initial structure already represents an energy-minimized and equilibrated structure. However, K527A, Y736A, and the double mutant K527A_Y736A move ~10 nm with sd ±2 nm along this Eigenvector, indicating a major conformational change with respect to the initial crystal structure. Like WT HsdR, mutant D796A does not show any change in projection along the first Eigenvector and oscillates around zero.

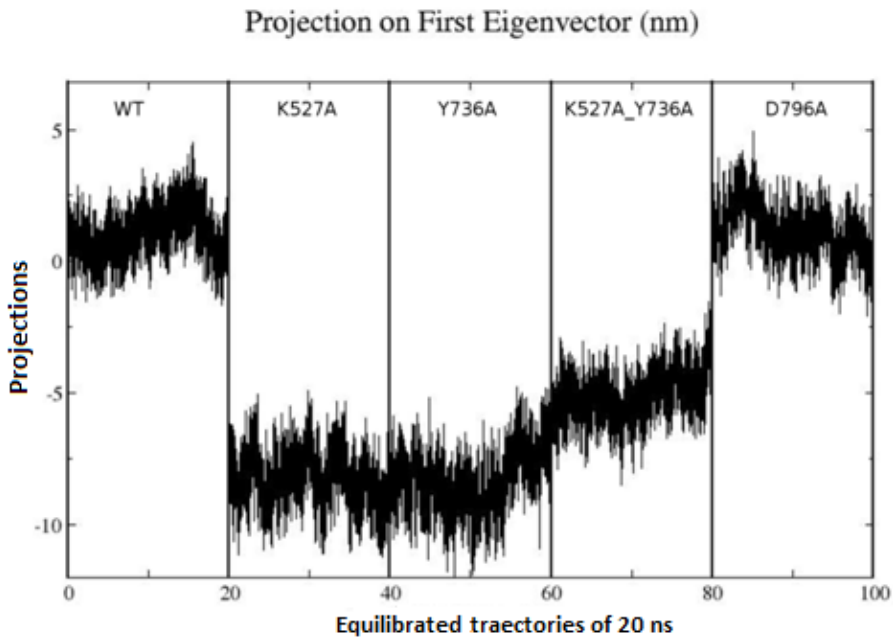


Fig. 3.2.3: Projection of the first Eigenvector of the joint trajectories onto the equilibrated part of each individual simulation. Data for WT simulations run from 0 to 20 ns, for K527A from 20 to 40 ns, for Y736A from 40-60 ns, for K527A_Y736A from 60 to 80 ns and for D796A from 80 to 100 ns.

To demonstrate global structural changes in helicase 2 domain during the simulation, root mean square deviations (rmsd) have been calculated for C-alpha atoms by superimposing the last snapshot of simulated structure

of WT and mutant on WT crystal structure. We calculated rmsd values that depict the inter-domain movement (due to difference between crystal and simulated structure) of helicase2 with respect to helicase1, where intra-domain structural changes have been filtered out of the final value of rmsd. It has been shown that during the simulation of 100 ns the helicase2 domain maintains an almost identical structure to that of the crystal. In the Y736A mutant the maximum rmsd is about 4.1 Å and as the 3.8 Å deviation in the K527A mutant simulation it indicates significant structural changes occurring during the simulation. The double mutant has a little bit less, 2.8 Å, while the D796A single mutant simulation shows a minimum rmsd of 0.6 Å with respect to crystal structure. Root mean square deviation analysis reveals that in WT and in the D796A mutant simulation, the helicase2 domain maintains the structural orientation as in WT crystal structure but in point mutation K527A, Y736A and double mutant K527A_Y736A, there is substantial structural changes occur during the simulation which affect the opening of the cleft.

Interaction analysis

The persistence of the inter-domain contacts observed in the crystal structure was analyzed in the equilibrated part of the simulations. Panel A in figure 3.2.4 reveals that in WT the hydrogen bond between the amino group of K527 and the backbone O of N794 (black line) is lost completely after the first 20 ns, while the contact between the amino group of K527 and the carboxylate group of D796 is maintained after the initial equilibration time (Figure 3.2.4, panel A, red line) The helix-turn-helix domains connecting region between the hydroxyl group of Y736

and carboxylate group of E730 is maintained throughout the whole simulation in WT (**Fig. 3.2.4,** panel B). The nature of inter-domain interactions observed in WT are practically identical in mutant D796A, predicting a similar behavior as WT for this single-point mutation. Given the fact that the contact involving the backbone contact of K527 is not persistent throughout the simulation while the D796 contact in WT is, it is surprising that in mutant D796A the contact between backbone O of N794 and the amino group of K527 is persistent (**Fig. 3.2.4**, panel C), and also the interaction between the hydroxyl group of Y736 and the carboxylate group of E730 is maintained throughout the whole simulation (**Fig. 3.2.4**, panel D). Thus Mutant K527A is unable to hydrogen bond with either N794 or D796, however the helix-turn-helix domains connecting regionwithY736 and E730 is persistent throughout the simulation (data not shown). In the simulation of mutant Y736A the interaction of K527 and D796 is lost during equilibration and the contact between amino group of K527 and backbone O of N794 is persistent (figure not shown).

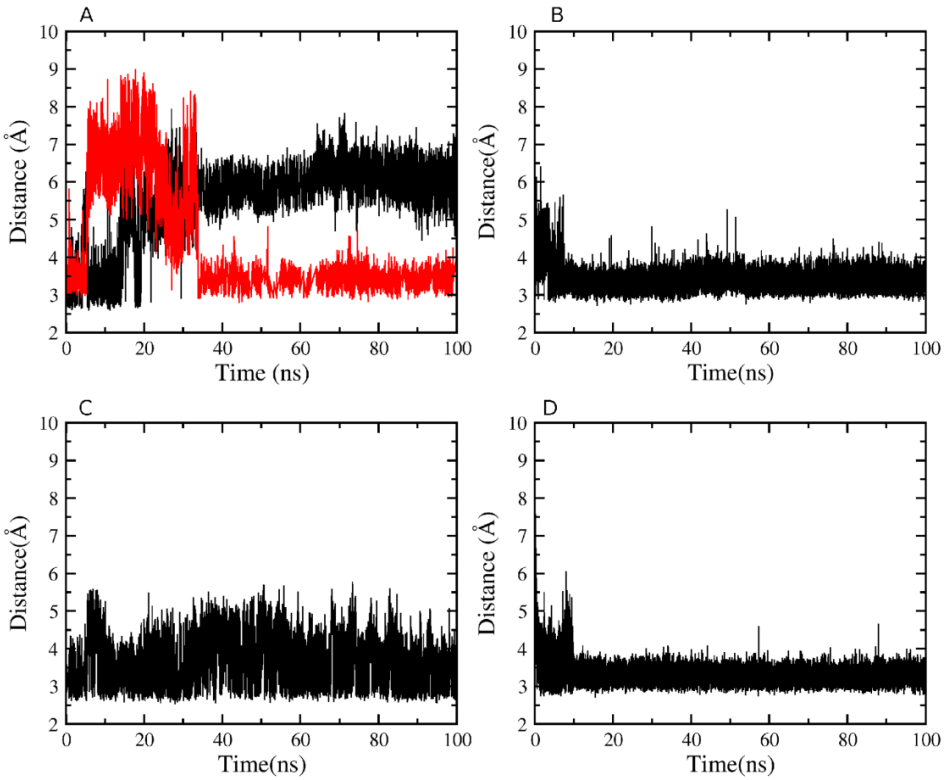


Fig.3.2.4: Contact distance in the course of the simulations. (A) WT K527(Nz)-N794(O) in red, WT K527(Nz)-D796CG in black (B) WT Y736(OH)-E730CD (C) D796A K527Nz-N794(O) (D) D796A Y736(OH)-E730CD.

The opening of the cleft is quantified by the average number of H-bond between the helicase 1 and the helicase 2 domains during the simulations as given in **Table 3.3.1**. In WT the number of inter domain hydrogen bond interaction is 13 ± 1 which keep the inter-domain distance 29.3 ±0.14 Å from the center of mass. In mutants K527A, Y736A, and K527A_Y736A the cleft opening corresponds to lower number of hydrogen bond interaction and drops to 9-10. In the simulation of mutant D796A, increase of number of hydrogen bonds to 17 ± 2 eventually decrease the distance by 29.01 ± 0.13 Å between the two domains and

the helicase domains get closer to each other. Our inter-domain H-bond and inter-domain distance analysis indicates that residue which is responsible for helicase2 and helicase interface interaction, play an important role in keeping the orientation of helicase 2 domain and by mutating these residue the complex is unable to maintain the viable conformation.

Table 3.3.1: Inter-domain distances and H-bonding between helicase I and helicase II domains

	Avg. Inter-domain Distance(Å)	Avg. Number of Interactions (H-bond)
WT	29.30 ±0.14	13 ± 1.40
K527A	29.77 ±0.18	10 ± 1.05
Y736A	29.90 ±0.17	09 ± 1.17
K527A_Y736A	30.30±0.21	09 ± 1.05
D796A	29.01 ±0.13	17 ± 2.11

Conclusion

The analysis of the hydrogen bond network suggests that in case of the mutant D796A the H-bond between K527 and N794 is sufficient to maintain the fold, and no major global motion is observed in the principal components analysis. K527A, Y736A, and double K527A_Y736A mutations at the helical-helicase 2 domain interface lead to major structural changes, and PCA analysis identifies the largest global motion to be an opening of the helicase cleft that gets wider relative to WT. In summation, the biochemical characterization, structural analysis and molecular dynamics simulations suggest that the second RecA like domain needs very specific contacts with the fourth domain to sample all stages of the translocation cycle.

3.3 The role of motif III and its extended region in positioning the two helicase domains in the motor subunit of the restriction–modification system EcoR124I

Introduction

The two helicses I the crystal structure of the hsdR motor subunit of EcoR124I are clustered with the characteristic motifs.Motifs I and II, also known as Walker A and B, interacts with α and β phosphates of the ATP. Motifs Ia, Ib, IV and V are involved in nucleic acid binding. Motifs III, V and VI (Arginine fingers) are reported to be required for coordination of α, β, and γ phosphates of ATP (Tanner & Linder, 2001; Caruthers & McKay, 2002; Cordin *et al.* 2006; Bleichert & Baserga, 2007). However, in the crystal structure of EcoR124I (Lapkouski, 2009) the closest distance between γ phosphate and the C-alpha of the conserved glycine residue is 7.8 Å and at least in the crystallized conformation a direct involvement of the conserved GTP sequence of motif III in phosphate coordination can be excluded.

In two recent studies of the closely related yeast SF2 translocase Rad54 (Burgess, 2013; Zhang, 2013) mutational analysis of different conserved amino acid residues from the extended motif III sequence revealed their importance for ATPase activity. Substitution of two highly conserved aromatic residues Y494 and F495 to alanine resulted in a dramatic reduction of ATPase activity. In contrast to WT Rad54 this double mutant has been shown to be defective in branch migration of a Hollyday junction (Kreici, 2013; Zhang, 2013). While these two residues do not have an equivalent in EcoR124I, mutations in the conserved positions

488 and 491, Q488A and L491Q, lead to a drastic reduction in ATPase and D-loop formation activities (Burgess, 2013). Docking of a dsDNA filament into the Rad54 structure, modeled from the X-ray crystal structure of zebrafish (PDB ID: 1Z3I), suggests that extended motif III in Rad54 directly contacts DNA. Thus the extended region of motif III is important for dsDNA binding which is likely linked to ATPase activity for Rad54 (Zhang, 2013), while a mutation within the core motif III G484R demonstrated the conserved glycine to be important only for DNA-dependent ATPase activity but not for dsDNA binding (Smirnova, 2004).

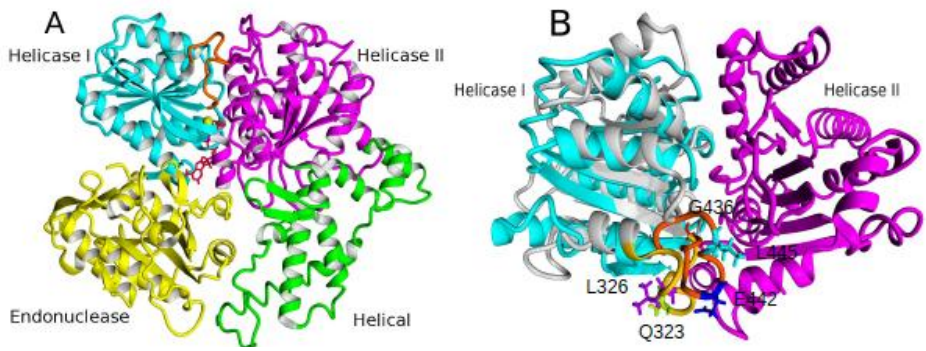


Fig.3.3.1: Three dimensional arrangement of the extended region of motif III. Structure of HsdR consisting of four domains which are color coded here, Yellow is endonuclease domain while Helicase 1 and Helicase 2 domains are depicted cyan and magenta, respectively. The helical domain is shown in green. ATP is represented in red color as a skeletal model and magnesium is indicated by yellow sphere. The extended motif III is represented by orange color. B) shows structural alignment of HsdR of EcoR124I from *E. coli* with Rad54 from *Danio rerio*.The crystal structure of HsdR of EcoR124I (2w00) and Rad54 from *D. rerio* (Zebrafish, PDB ID: 1Z3I) were used for structural pairwise alignment in Yasara using helicase 1 domains (in cyan), RMSD of Cα is 2.352 Å and helicase II domain (in magenta). Residue Q448 and L491 in Rad54 from *Saccharomyces cerevisiae* correspond to Q323 (light green) and L326 (purple) of Rad54 from *D. rerio*, in side chain respectively. Top view, HsdR in gray, Rad54 in grey, HsdR loop representing the extended motif III and containing amino acid residues for mutagenesis is shown in orange and the

corresponding loop from Rad54 in dark yellow. Helicase domain 1 is on the left and helicase domain 2 is on the right, the respective loop in each protein is situated between both helicase domains. Residue. G436 (red),E442 (blue)and L445 (cyan) are represented by side chain of HsdR. C) shows the conformational alignment of the HsdR loop in orange and the corresponding loop in Rad54 in yellow while the respective residue are shown in side chain keeping the above color representation.

The reported critical residues in this extended region of motif III are partially conserved in HsdR of EcoR124I (residues 436-448), and therefore indicateda certain possibility that an extended motif III might be required for translocation by EcoR124I, too. Mutations in two critical positions, Q448 and L491, show a loss of ATPase activity in case of Rad54 (Burgess, 2013). In HsdR of EcoR124I the corresponding residues are Glu442 and Leu445 and are positioned on a structurally conserved loop close to the interface between helicase 1 and 2 domains, and thus in a similar environment as in the Rad54 protein. Corresponding point mutants of EcoR124I have been prepared to thoroughly tested *in silico*, *in vitro* and *in vivo* toelucidate the role of the extended motif III. Here we show that point mutations in the extended region have minor or slight effect on the endonuclease activity, ATPase activity and that the restriction phenotype is preserved as in WT. Additionally, a single-point mutation of the conserved glycine to alanine in motif III was prepared. For the case in which a slightly larger side chain in position 436 was introduced (Ala instead of Gly)molecular dynamics simulations show that the experimentally observed impairing in ATPase activity/translocation and decrease in consequent DNA cleavage,can be explained by an increase in the interdomain distance provoked by a slight decrease in interdomain hydrogen bonding resulting in a widening of the cleft between both helicase domains.

Results

Global conformational change:

Global motions during the simulations were analyzed by principal component analysis (PCA). Therefore always the last 20 ns of each of the five individual trajectories were joined into one long trajectory to allow the calculation of identical Eigenvectors for each simulation by running the PCA over the combined trajectory. This results in a single set of eigenvectors that apply to all equilibrated trajectories.

The first eigenvector describes the main movement seen in the simulations, which is a collective motions of the helicase 2 domain with respect to helicase 1. **Fig.3.3.2** shows the projections of the first Eigenvector on all the system. In the G436A simulation a maximum value for the eigenvector is observed that ranges from -18 to -10 while the average movement along this eigenvector oscillates between 0-5 nm in WT and the other mutants.

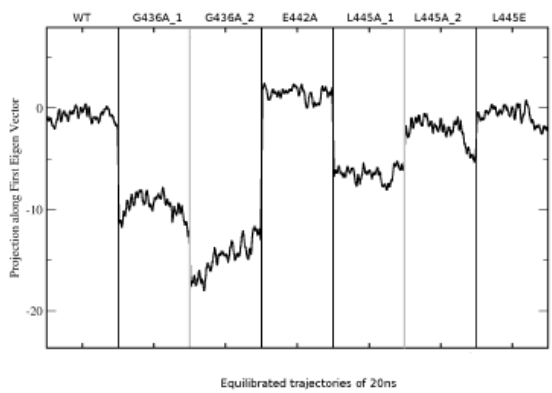


Fig. 3.3.2: **Principal component analysis.** Projection of the first eigenvector onto the joint trajectories of equilibrated 20 ns of each individual simulation of WT and mutants. Repetitions for G436A and L445A simulations show variation in projections while projections for WT, E442A and L445E simulations are quite same in repetitions.

An overlay of the structures before and after the simulation in the **Fig. 3.3.3** and **3.3.5** reveal the displacement of helicase domain 2 in mutant G436A. This displacement is reflected in an increase of the distance between the center of mass of helicase 1 and helicase 2 domains. While the average distance between helicase 1 and helicase 2 domain is around 29 Å in WT and all other mutants except, the average distance between the domains increases to 30–31.2 Å in G436A. This global collective motion can be described as an opening of the cleft between both domains as shown in **Fig. 3.3.5**.

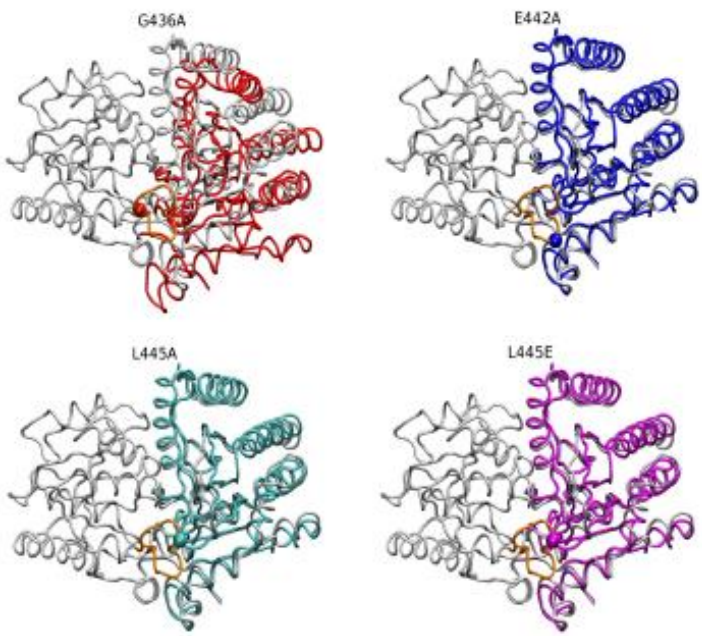


Fig.3.3.3: Overlay of the crystal structure of HsdR with simulated WT structures from equilibrated trajectories. Both domains of the translocase crystal structure (helicase 1 and helicase 2 domains) are shown in tube representation in gray. The helicase 2 domain of WT is in orange. Conformations were fitted on the helicase 1 domain of the crystal structure.

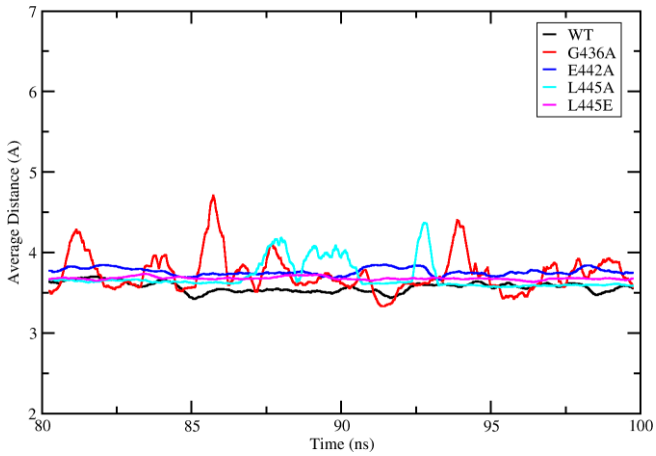


Fig.3.3.4: Distance plot for the most persistent interaction between the helicase domains. The average distance between Cγ of ASP341 in the helicase 1 domain and Oγ1 of THR661 in the helicase 2 domain of the last 20 ns of trajectories. Black color indicates WT, red G436A, blue E442A , cyan L445A and magenta is L445E.

To investigate the relationship between the observed essential dynamics and residue interactions, the persistence of interdomain contacts during the simulations was analyzed. The most significant contact between the two domains is a hydrogen bond between residue D341 in helicase domain 1 and residue T661 in the helicase 2 domain. While in WT and the other mutants this contact is persistent throughout the simulations as shown in Fig.**3.3.4**, in G436A this contact is less persistent and lost frequently in the second half of the simulation.

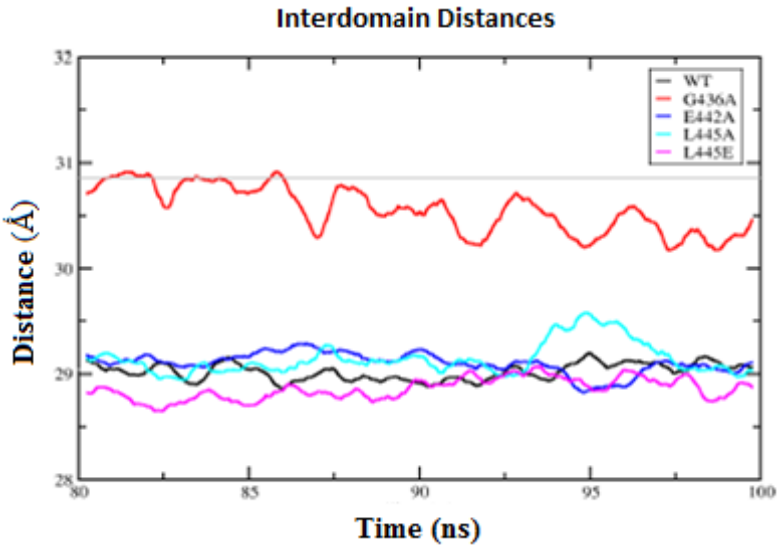


Fig.3.3.5: Interdomain distances. Average interdomain distance between the centre of mass of helicase 1 domain and the center of mass of helicase 2 domain in the last 20 ns of the equilibrated trajectories. Black color indicates WT, red G436A, blue E442A , cyan L445A and magenta is L445E while the horizontal line in maroon represent the interdomain distance of helicase 1 and helicase 2 in the crsystal structure of NS3 RNA helicase (PDB ID: 1A1V) and horizontal line in brown represent the inter domain distance of helicase 1 and helicase 2 from the centre of mass in the crystal strucure of HsdR of EcoR1241 (PDB ID: 2W00).

Hydrogen bonding plays a crucial role in stabilizing domain-domain interfaces. Hydrogen bonds between helicase 1 and helicase 2 domain were quantified based on the two criteria: (a proton donor (D) and acceptor (A) distance of 3.5 Å or less and (2) – D-H ... An angle of 120 degrees or more. WT and all other mutants except G436A show a well conserved number of 12-14 hydrogen bonds between helicase 1 and helicase 2 while in G436A a lower number of 10-12 hydrogen bonds is observed during the last 20 ns of the equilibrated trajectories, when the cleft between the domains is more open. Taking into account all the simulations the analysis of H-bond reveals that these interaction are

responsible for the proper conformation of the translocase domains which leads to the ATPase activity.

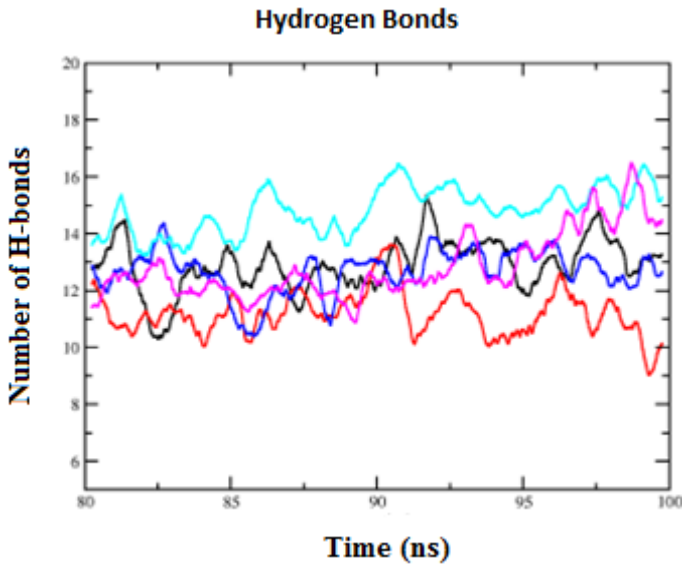


Fig.3.3.6: Hydrogen bond Analysis. Average number of H-bond between Helicase 1 and Helicase 2 in the last 20 ns of equilibrated trajcetories. WT simulation is represented in black, G436A in red,E442A in blue,L445A in cyan and L445E in magenta color.

Overlay with known conformation

Several crystal structures of helicases have been solved with ssDNA and with various ATP analogs and these structures represents snap shots of different conformations of the helicases in the translocation cycle along single stranded DNA. Internal conformational motions driving by binding of ATP molecule, its hydrolysis into adenosine diphosphate (ADP) and a phosphate Pi, generate forces on the nucleic acid strand and allow the helicase to translocate along the DNA. Analyzing the reported crystal structures of SF2 helicases in various conformations we found that, these conformational changes leads to either variation of distances

between the domains or certain degree of rotation occurs in the helicase 2 domain with respect to the helicase 1 to complete the translocation cycle. During the simulation of mutant G436A the global conformational change in the helicase 2 domain is quite significant with respect to the crystal structure as shown in fig.3.3.3, while in simulation of WT and in other mutants these conformational change in the helicase 2 domain is quite small. We overlay the conformation of the helicase 2 with extreme projection achieved in last 20 ns to the known crystal structure of SF2 of helicases by fitting on the helicase 1 domain, we found that the closest conformation of the helicase 2 domain from the G436A mutant simulation was with the structure of hepatitis C virus NS3 RNA helicase (PDB ID: 1A1V) complexed with ssDNA oligonulceotide (**Fig. 3.3.7**). This crystal structure also contains seven highly conserved motif to represent SF2 helicases which play a key role in nucleotide triphosphate binding/hydrolysis and nucleic acid binding. The protein consist of three structural domains with the oligonucleotide lying in a grove between the first two domains and the third. Domain 2 is flexibly connected to domain 3 and shows the rotation with respect to domain 1. This DNA bound structure also leads to the proposal of a mechanism of ATP dependent DNA translocation that is consistent with the enzymatic and structural properties of other helicases. The whole topololgy of the domains are quite similar with the conformation achieved in G436A mutant simulation in last 20 ns of equilibrated trajectory. Thus this conformation of G436A is also viable conformation and may represent one of the conformation of the translocation cycle.

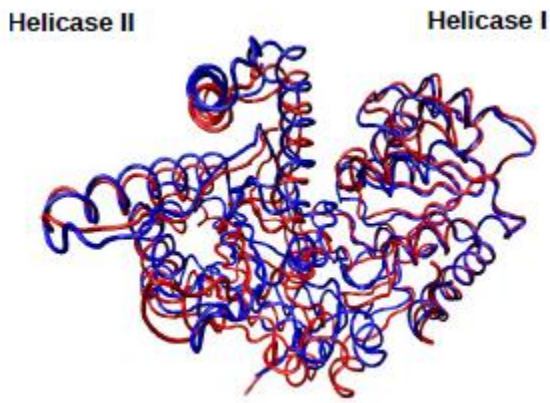


Fig.3.3.7: Overlay of the helicase domains from the crystal structure of NS3 and simulated G436A. Snapshot structure from equilibrated region of simulation of G436A. Helicase 1 and helicase 2 domains are shown in tube representation. NS3 crystal structure in lemon, G436A in red. Mutant conformation was fitted on helicase 1 domain.

Conclusion

The helicase motif III in Rad54 has an extension which is essential for ATPase activity and the similar structural environment as well as the conservation of the residues predicts a possibility that the extended motif in HsdR might contact dsDNA and participate in translocation, too. However, the extended region of motif III, although it is similar to Rad54 in 3D as well as in sequence, does not significantly influence ATPase activity or restriction activity in EcoR124I. Thus in contrast to Rad54, the extended motif in EcoR124I does not seem to participate in contacting dsDNA.

The functional consequence of G436A in motif III is an alteration of the ATPase activity/translocation which is then promoted to restriction activity and molecular dynamics simulations show that the introduction

of a slightly larger side chain in position 436 leads to a slight decrease in interdomain hydrogen bonding and a consequent increase in the interdomain distance. These structural changes result in a large conformation change, a widening of the cleft between both helicase domains. The resulting conformation does not necessarily represent a non-functional or previously not existing conformation but might well represent another translocation stage of the RecA-like helicase. The orientation of the helicase domains in the crystal structure of NS3 RNA helicase with ssDNA substrate and the conformation achieved in G436A mutant simulation are quite similar.

3.4 Alternative conformation of 180 loop

3.4.1 Introduction

In one of the crystal structure of the mutant (E165H) of HsdR we found an alternative conformation of the 180s loop. In this conformation ARG182 is neither in contact with 220 loop nor does it is in contact D881 but R182 is in contact with D855 and makes a salt bridge contact with the helical domain. This new conformation of the loop can either be a feature exclusively of the mutant enzyme that traps the loop in this state or regularly occur in WT, too, however not as the most populated state. To test whether salt-bridge interaction between Arg182 and Asp855 in the Glu165His mutant is the cause of reduction in ATPase/translocation activity reported earlier (Sisakova *et al.*, 2008), we introduced a second mutation Asp855Ala in the Glu165His mutant to nullify the possibility of a salt-bridge. As well we additionally substituted Arg858 with alanine in the Glu165His mutant to remove the stacking interaction stabilizing the Arg182-Asp855 salt bridge. Also single mutations Asp855Ala and Arg858Ala were introduced. With the help of molecular dynamics simulation we explore the alternative conformation where the 180s loops interacts with D855 and might have a similar function as D881 in WT. We were also interested to see the conformational changes of the endonuclease domain that we reported earlier for WT (Sinha *et al.* 2014).

3.4.2 Results

As shown in our previous work (Sinha .et.al) the endunuclease domain is regularly in contact with LYS220 and this contact contributesto ATP

binding similar as do other contacts in theATP binding pocket. During the simulation of WT this contacts breaks and allows the interaction of the 220s loop with the 180s loop via an interaction with ASN221 and ARG182. The distance between O atom of ASN221 was in H-bond distance with the side chain of ARG182 and this interaction was persistent through the whole trajectory after initial equilibration period as shown in figure 2 (black line) and this interaction was also pretty stable in the R858A simulation. In the D855A simulation initially the contact between the 220s loop and 180s loop persisted but after equilbration of about 50 ns this interaction gradually broke and after 80 ns it was completely lost and ARG182 switched to another conformation and approached the helical domain. Surprisingly, TYR183, which is part of the characteristic QxxxY motif, made a stacking interaction as shown in Fig.3.4.3and3.4.4 with ARG858, while in the R858A simulation R182 kept the contact with 220s loop and TYR183 also did not find any counterpart and finally flipped towards the 220s loop, **Fig.3.4.2**.

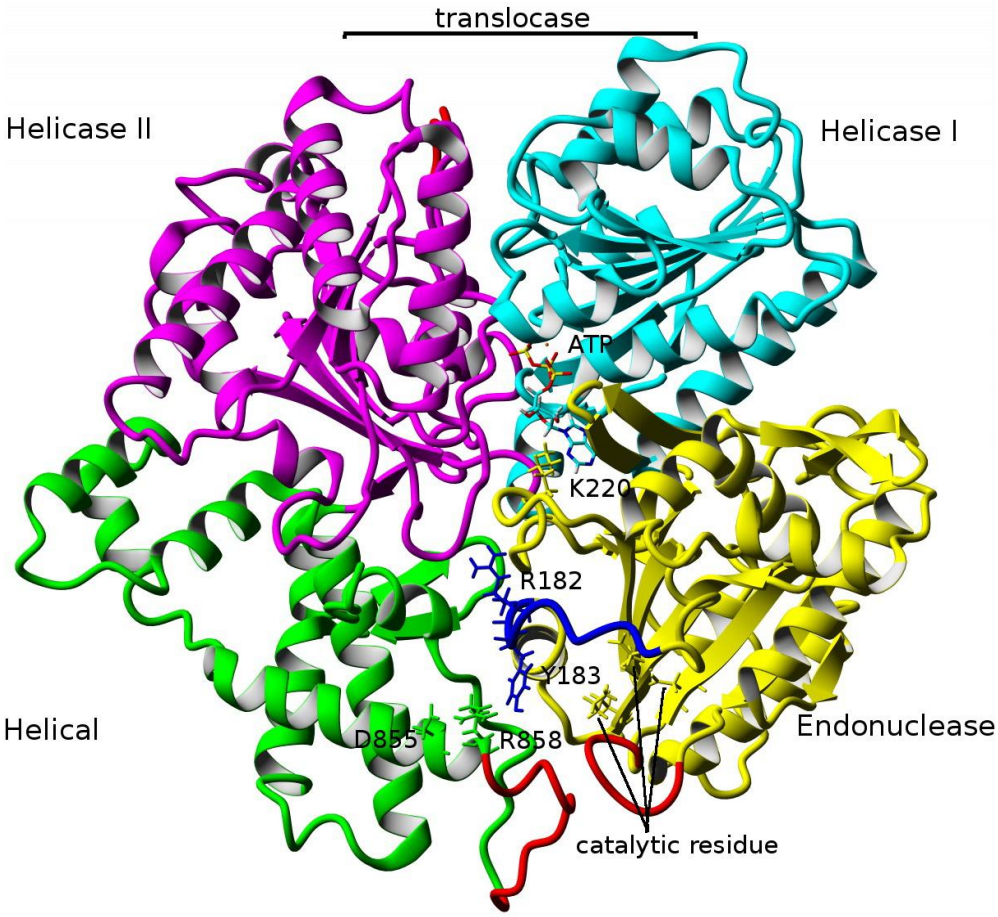


Fig.3.4.1: Crystal structure of the motor/HsdR subunit.Selected side chains relevant for this work is also shown and QxxxY motif is shown in red color. Other residue of interest is shown by its sidechain.

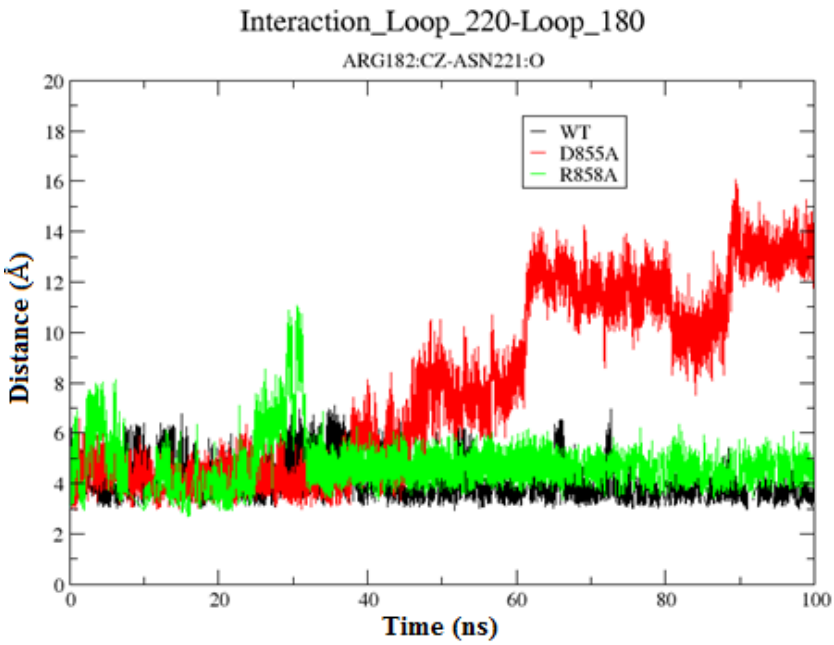


Fig. 3.4.2: Representation of the interaction of 220 loop and 180 loop mediated by ARG182-ASN221.

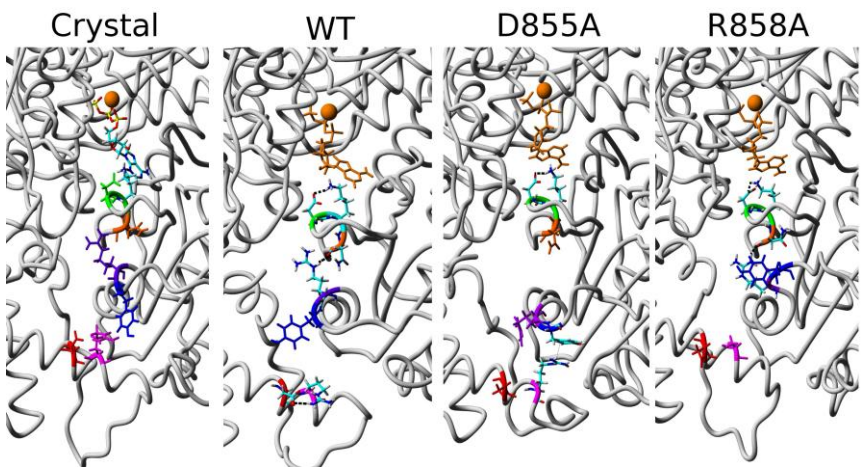


Fig. 3.4.2: (a) shows the contact of K220 to ATP (b) while during simulation the conatct breaks and new interaction found between R182 and N221 (c) R182 is in alternative conformation and during simulation stacking interaction found Y183 and R858 (d) in the muatnt simulation R858A where there is no interaction R182 is approaching back towards 220 loop.

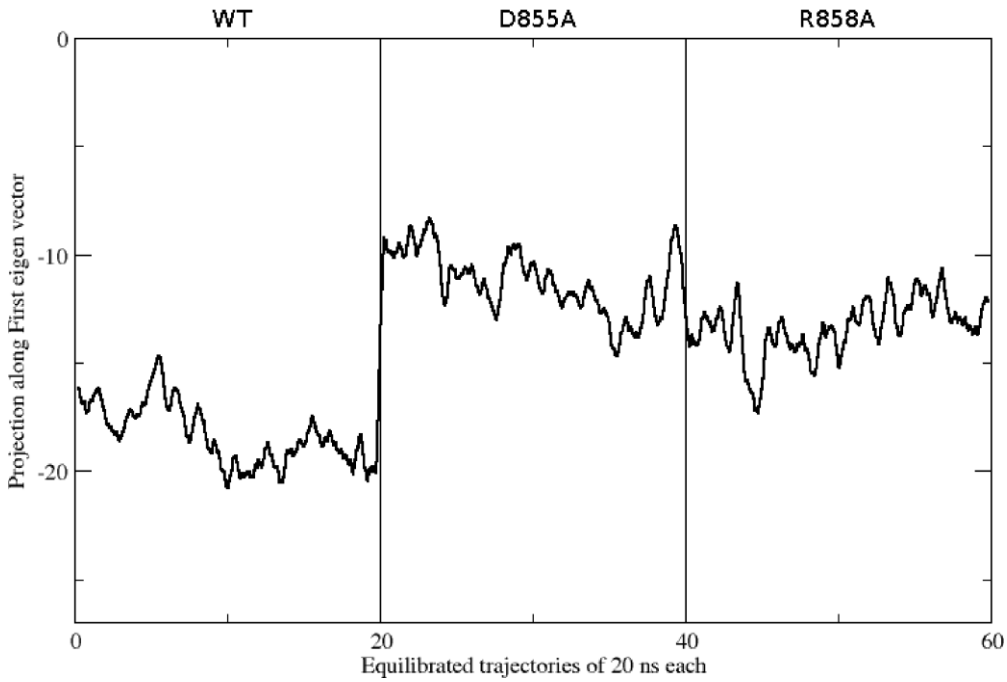


Fig. 3.4.4: Projection of endonuclease domain with respect to rest of the structure along the first Eigen vector in WT and mutants.

To quantify the global conformational changes in the endonuclese domain PCA have been done which represent the motion of endonuclease domain with respect to rest of the protein. To represent collective motion along the first eigen vector,trajectories of the last 20 ns of every simulation were joined. Running the PCA over this concatenated trajectories results in a single set of eigen vectors that apply to all equilibrated trajectories.

The biggest motion during the simulations is described by first eigen vectors and **Fig.3.4.4** shows the projections of every simulation in last 20 ns along the first eigen vector. In WT simulation the projection is maximum and it is around 20 nm with respect to crystal structure which is taken as reference and represented by the 0 value.The projection is

much less in the D855A simulation but in the R858A simulation we found a significant difference with respect to crystal structure.

An overlay of the structures is shown in **Fig.3.4.5** where the gray color represents the initial structure and for clarity in the representation only helical and endonuclease domains of the crystal structure are shown. We observe that there is significant motion in WT (blue) as well as R858A mutant (green) while D855A mutant (red) shows no or almost negligible motion.

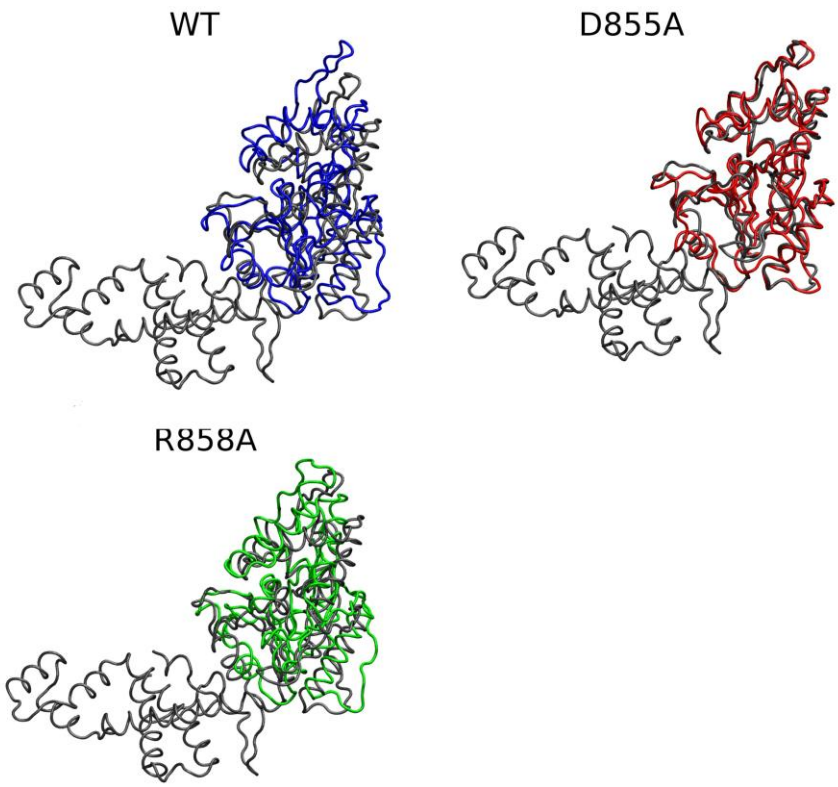


Fig.3.4.5: Overlay on crystal structure of the WT endouclease domain and the mutant simulated structure.

Conclusions

In MD simulations the new conformation is rarely sampled in WT and R858A. In mutant D855A it is sampled and showed specific interactions which suggests that this conformation is trapped in the E165H mutant andmight be only sampled under very specific circumstances.

3.5 Involvement of the conservative domains of the DNA recognition subunit HsdS in assembly of functional EcoR124I endonuclease.

3.5.1 Introduction

As previously described, the S-subunit plays a key role in DNA recognition as well as subunit interaction with the HsdM and HsdR subunits to form the complete pentameric functional enzymes. This subunit consists of a central conserved region and two highly variable regions known as target recognition domain (TRD). It has been reported that the central conserved region is responsible for the interaction with HsdM and HsdR subunit, although the interaction with HsdR might be a rather indirect one as a direct interaction was not predicted so far by any of the existing structural models.

Our collaborator Marie Weiserova and her team recently observed that point mutations in the border region of the central conserved part have a significant effect on the activity of the complex. To observe structural and conformational changes resulting from these point mutations we first model the structure based upon other known crystal structure of HsdS subunit and performed classical molecular dynamics simulations.

3.5.2 Model Building

Sequences of HsdS subunit of EcoR124I in FASTA format were retrieved from the Uniprot database (accession number P10485) and after running 3 PSI-Blast iterations, possible templates for model building were identified. According to the E-value (expect value), the closest crystal structure of HsdS from *Methanocaldococcus jannaschii*

(HsdS MjaXIP, PDB ID: 1YF2 having resolution of 2.4 Å) was selected as a template for initial model generation with an alignment score of 182.0 and sequence coverage of 93%. The overall sequence identity between target and template is 22.9% while the sequence similarity is 42.6%. In this structure 345 of 404 target residues (85.4%) are aligned to template residues (**Fig.3.5.1**). A set of 5 initial homology models was constructed by using fragment based and loop remodeling approaches in YASARA. 15 loops were modeled and optimized after the modeling of backbone structure. The generated model was further optimized by a combined steepest descent and simulated annealing method followed by a full unrestrained simulated annealing minimization to achieve the most stable conformation. The best homology model was selected according to structural quality parameter of the Z-scoring function. The final model (**Fig.3.5.2**) with the highest rank was chosen and had the overall Z-score of −1.541.

Under the model validation protocol, YASARA uses WHAT CHECK to validate the obtained model by generating a plot of the overall quality Z-score. The model quality of the residues in this plot was quite reasonable. The final model was also cross-validated by analyzing its Ramachandran plot (**Fig.3.5.3**). In the Ramachandran plot around 381 (94.8%) residues were in favored regions, 18 (4.5%) residues in allowed region, and 3 (0.7%) residues in outlier regions. The model was also evaluated by Verify3D. The score (≥ 0.2) averaged for 81.19% of the residues, suggested that the resultant model contains significant errors only in loop regions. This model of HsdS of EcoR124I has enabled the prediction of a number of functionally important residues potentially involved in DNA binding.

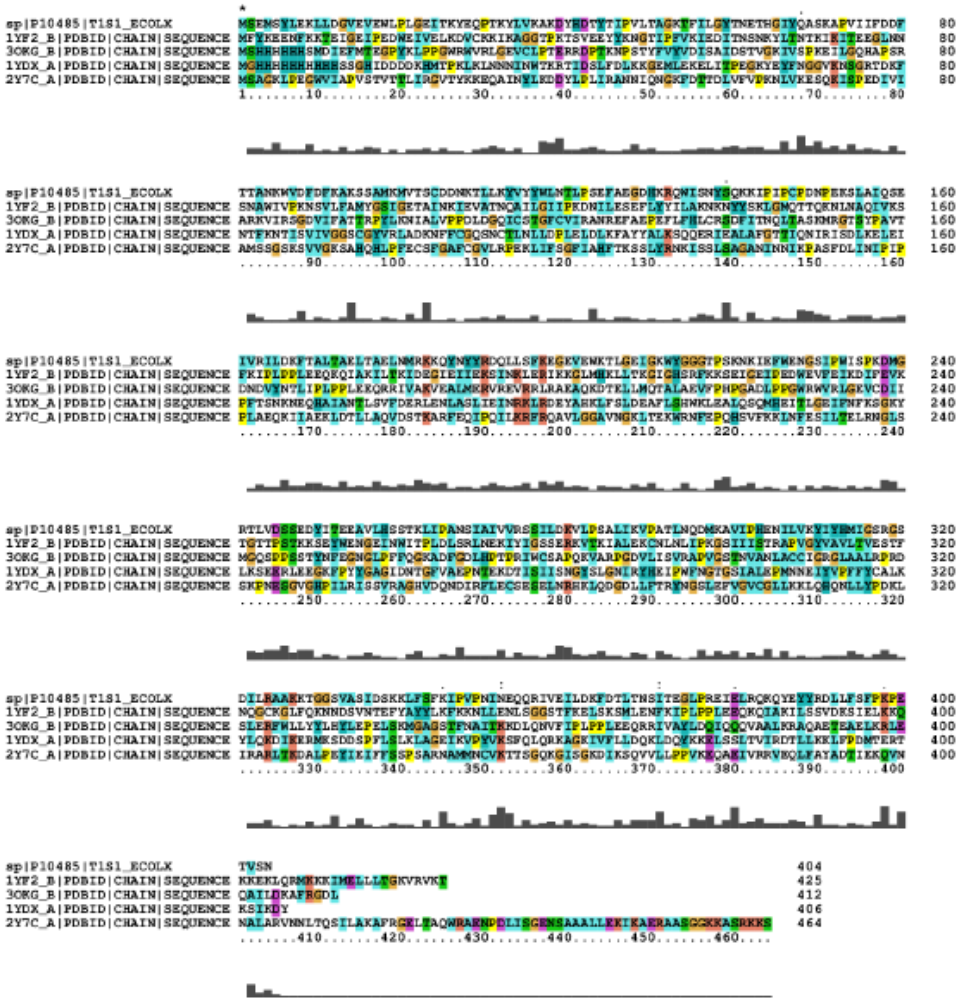


Fig. 3.5.1: Sequence alignment of the EcoR124I HsdS sequence with the templates used for modeling (crystal structure of HsdS from *Methanocaldococcus jannaschii* (HsdS MjaXIP, PDB code-1YF2 chain B having resolution of 2.40 Å) is the selected one for modeling). Color residue coding (orange for amino acids-GPST, red for HKR, blue for FWY, and green for ILMV). * represents the identical residues, colon represents similar residues, and dot represents different residues. The height of the histogram stands for conservation of the residue in the alignment.

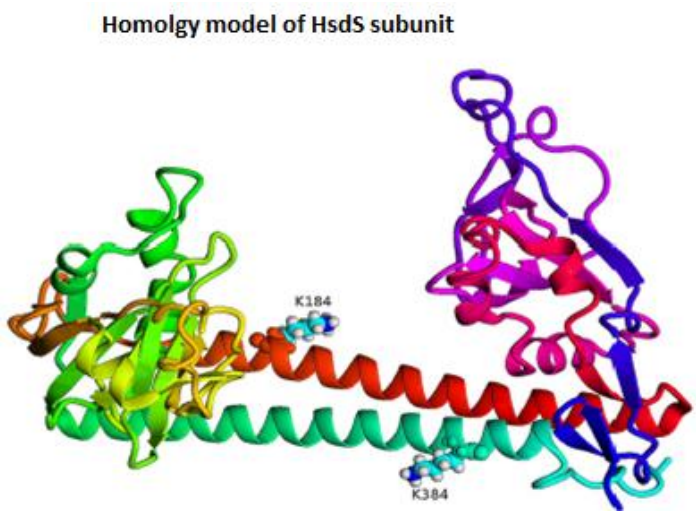


Fig. 3.5.2: Rainbow view of Homology model of HsdS of EcoR124I from N to C terminal in cartoon representation. Ball representaion of residues Lys-184 and Lys-384. Figure prepared in YASARA.

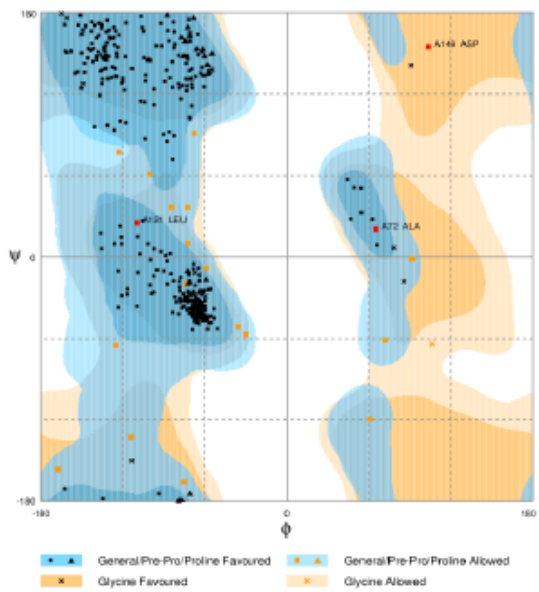


Fig.3.5.3: Ramachandran plot made by Rampage online tool (Lovell *et al.* 2002). Around 381 (94.8%) residues lies in favoured region, 18 (4.5%) residues in allowed region, and 3 (0.7%) residues in outlier region.

3.5.3 Molecular Dynamics Simulations

Intially 100 ns molecular dynamics simulations have been performed for the complete HsdS subunit along with TRD regions and C-terminal tail region. During molecular dynamics simulations of wild ytype HsdS subunit the TRD (Target Recognition Domains) started approaching each other within few nanoseconds of simulation and at the end of the simulation the two TRDs were interacting significantly with each other (**Fig.3.5.4**). This nicely corresponds to the precipitation observed *in vitro* for these conditions results which showed that the HsdS subunit is hydrophobic and not stable in water, but tends to aggregate when it is not in contact with the HsdM subunit. To find the specific consequences of point mutations K184N and K384N on the dynamics and struture of the two conserved helical domains and to neglect the precipitaion effect of TRDs only helical domains with capped endings were selected for further all atom molecular dynamics simulations.

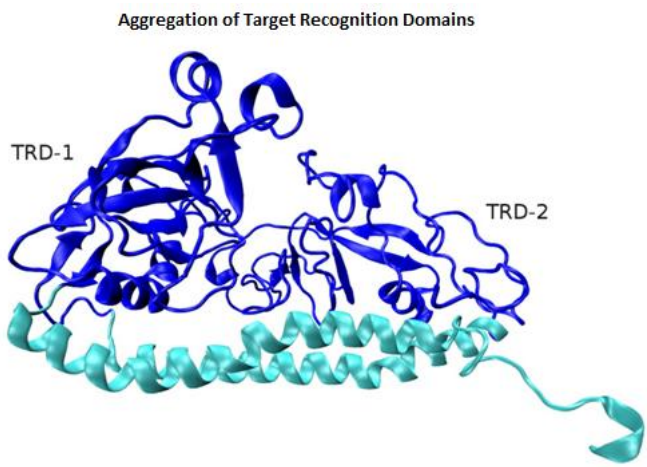


Fig. 3.5.4: Snapshot of simulation of HsdS subunit of EcoR124I showing aggregation of Target recognition domains (TRDs).

Visualization of the trajectories of wild type and mutants reveals that all the simulations are well equilibrated after 50 ns of simulations. Fluctuations in the individual residues during the simulations are measured by root mean square fluctuation (RMSF) plots. **Fig. 3.5.5** shows the fluctuations as an RMSF plot for the backbone for the equilibrated part of the simulations (last 20 ns). The average RMSF value for a typical secondary structure element is around 0.75 Å and ranges from 0.5 to 1.75 Å. To quantify the overall secondary structural changes throughout the simulations, DSSP plot were created that use Kapsch-Sanders parameres to identify secondary structure during a simulation. **Fig.3.5.6** shows that there is no change in secondary structure except at both termini of the helices.

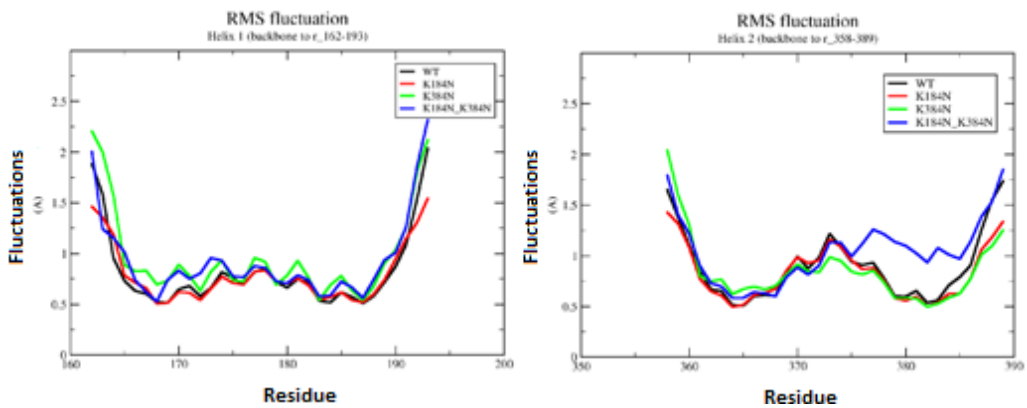


Fig.3.5.5: RMSF plot for Helix 1 (left) and Helix 2 (right) of backbone from equilibrated regions.

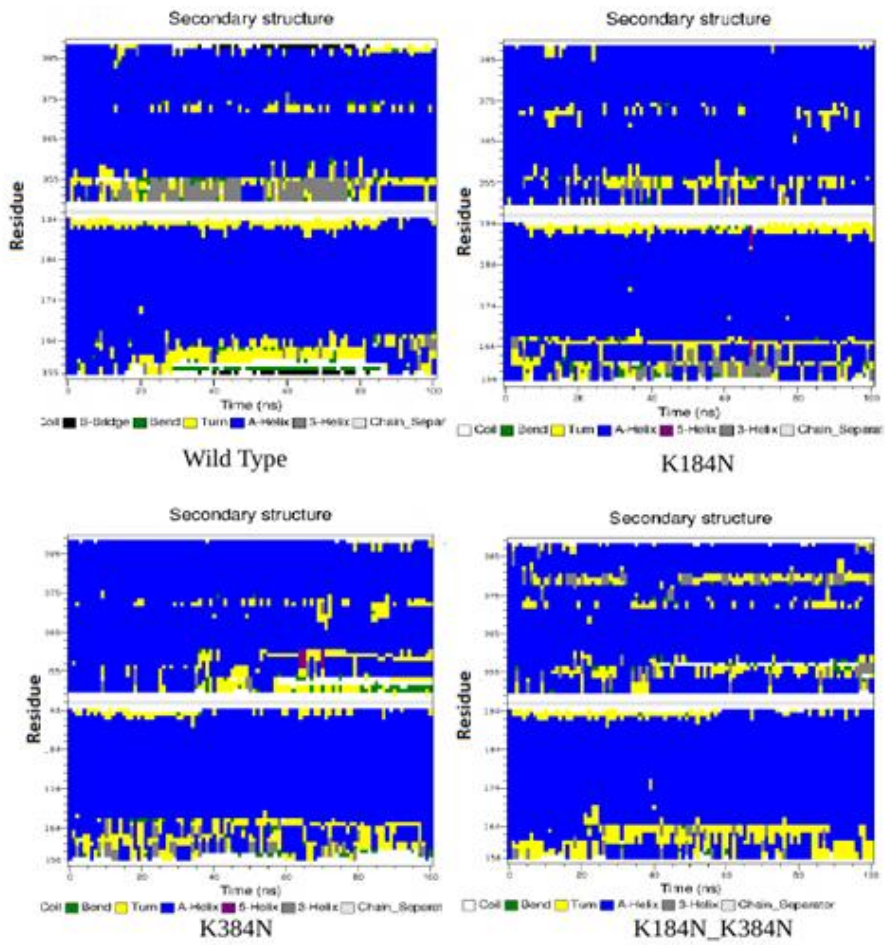


Fig. 3.5.6: DSSP plots for HsdS wild and mutant types show no structural changes neither in WT nor mutant simulations.

During the course of molecular dynamics simulation of wild type of 100 ns, kinking is observed in helix 2 bearing the K-384 residue while in the mutant K384N simulation, helix 2 becomes flatter, having relatively less kinking. On the other hand in K184N and double mutant K184N_K384N, helices become even more flat and rigid without having any kink formation. This kinking causes a change of orientation of the helices with respect to each other. To quantify this change, dial plot

analysis has been done using the Trajelix tool of the Simulaid package which measures global helix tilt angle on the x-axis.

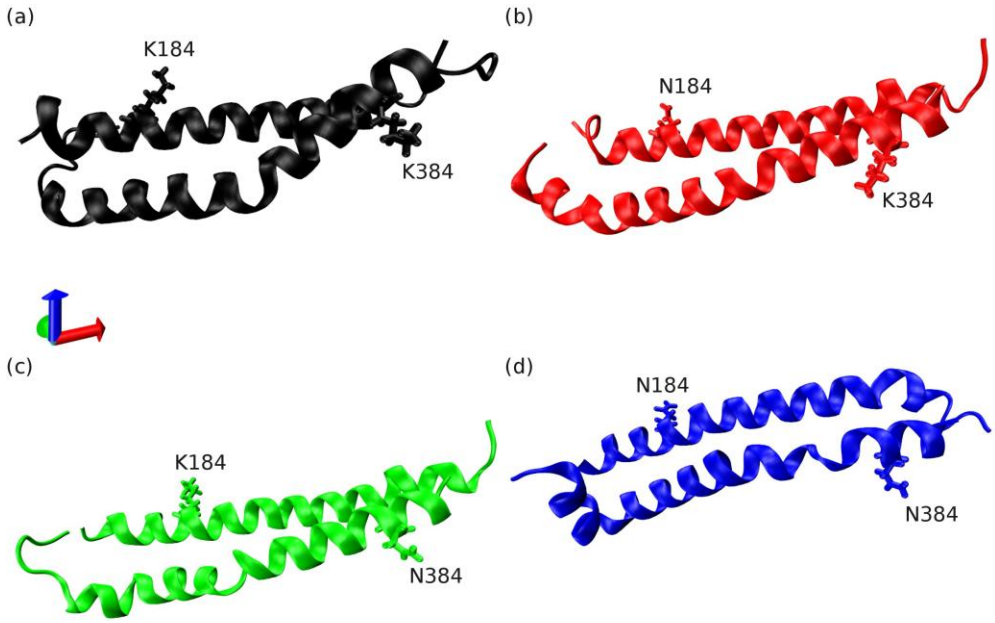


Fig. 3.5.7: Snapshots of helices of WT and mutant for 100 ns molecular dynamics simulations. WT is represented in black (a), K184N mutant in red (b), K384N in green (c), and K184_K384N in blue (d).

The dial-plot of wild type (WT) simulation demonstrates the highest tilt angle on x-axis in the helices with respect to each other. During the simulation the orientation of the helices changes to a twisted cross-over conformation as shown in **Fig.3.5.8**. The average twist angle between the two helices in WT simulation is 121.17 degrees. While the twist angle for K184N mutant is extremely small resulting in almost parallel/flat conformation. The twist angle is approx. 20 degrees and approx. 40 degrees in K384N and K184N_K384N mutant respectively.

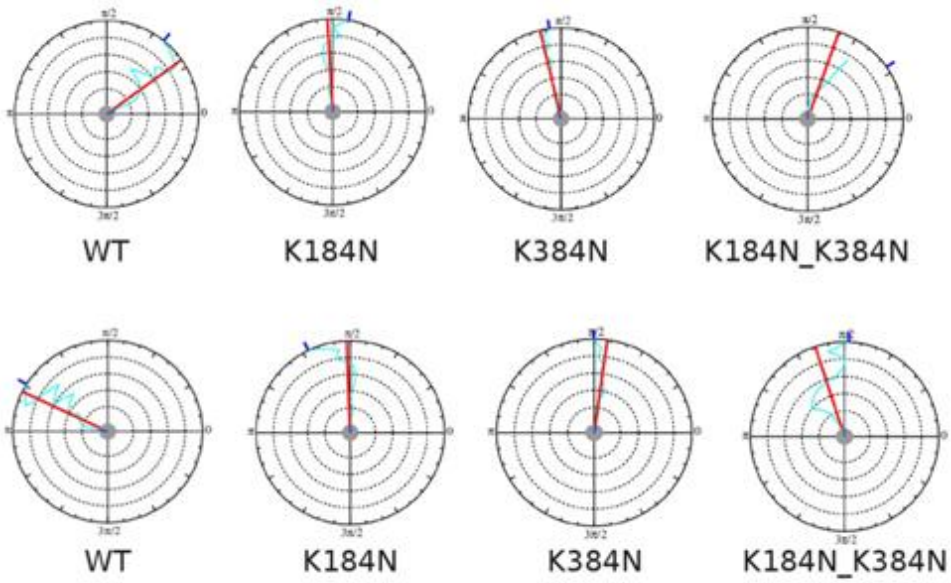


Fig. 3.5.8: Dial plots prepared by Trajelix plugin of Simulaid framework for the analysis of molecular dynamics simulations. Upper half of the figure represents the dial-plots for Helix-1 (K184N) and lower half is for Helix-2 (K384N). The evolution of a selected residue's Psi and Phi angle represented as dialplots. The axis of time is the radius of the plot. The line in the central smaller disk gives the initial value, the red line represents the average value and the green tick outside the whole disc shows the final value.

The above conformation achieved by helix-1 of WT simulation corresponds to highest solvent surface accessible area (SASA) of 45.7135 $Å^2$ which corresponds to higher penetration of solvent while the lowest SASA is 44.2877 $Å^2$ in Helix-1 of K184N mutant (**Table 3.4.1**). Comparison of SASA analysis shows that the conformation of helices achieved in mutant K184N simulation corresponds to less shielding from solvent exposure of its atoms than observed in WT helices whose orientations were cross-linked with each other with greater flexibility to make the most viable conformation.

Table 3.4.1: Solvent accessible surface area in $(Å)^2$ with its standard deviation.

	Helix-1	Helix-2
WT	457.135+/-10.9648	468.611+/-12.5147
K184N	442.877+/-10.2318	463.15+/-9.24241
K384N	456.668+/-9.8674	465.915+/-10.6417
K184N_K384N	446.3+/-8.5157	464.431+/-12.132

Conclusion

The S subunit is hydrophobic and insoluble in water. It tends to aggregate when it is not in contact with HsdM subunit. This is in full accordance with the precipitation observed *in vitro* for these conditions. The model of HsdSsubunit of EcoR124I has allowed to predict a number of functionally important residues potentially (indirectly) involved in DNA binding.The point mutations in the conserved domain of HsdS subunit in positions K184 and K384make the helices more rigid, probably trapping the HsdSsubunit in a conformation in which the methyltransferase complex cannot assemble properly with the HsdRsubunit, resulting in a loss ofrestriction activity.

4.Conclusions

Understanding of molecular interactions in particular physiological conditions is of great importance in biomolecular complexes. The major work of my PhD thesis is focused on the computational study of the relevance of inter and intra domain intractions in the functionality of motor/ HsdR subunit of the Type I restriction modification systemEcoR1241. At the end I would like to draw a few major conclusion -

1. In the crystal structure of HsdR subunit, endonuclease domain is unexpectedly contacting ATP via Lys220 and our QM/MM calculation found that the contribution of Lys220 in binding of ATP is comparable with other adenine binding residue of the binding pocket and excluding a by-chance contact or crystallization artefact.

2. Molecular dynamics simulations studies reveal an additional role of the QxxxY motif via Arg182 and explain the interdomain communication from ATP to the catalytic center across the enzyme about its ATP ligation status. We showed that substitution of the Arg182 sidechain with a shorter sidechain leads to an interruption in the communication of the translocation and endonuclease functions.

3. The observed rotation of the endonuclease domain with respect to the other three domain takes place along a screw axis that passes through center of mass of the structure and may convey information about the ATP ligation status of the enzyme complex.

4. The biochemical characterization, structural analysis and molecular dynamics simulations suggest that the absence of the helical-helicase 2 interface interactions would lead to a conformational shift towards the partial open conformations, while the more closed conformation is required for the directional motion of the helicase domain 2 to produce the ATP power stroke to advance the DNA.

5. The second RecA like domain needs very specific contacts with the fourth domain to sample all stages of the translocation cycle, and the fourth domain is thus directly and unexpected involved in translocation and endonuclease activities.

6. In EcoR124I this extended region, although it bears sequence and structural similarities with Rad54, does neither influence ATPase activity nor restriction activity**.**
7. The conserved glycine residue in motif III plays a role in the positioning of the two helicase domains towards each other and its mutation alters ATPase activity and cleavage activity.
8. Simulations with the alternative 180s loop conformation demonstrate that the new conformation is rarely sampled in WT and R858A suggesting that this conformation is trapped in E165H mutant and in WT might be only sampled under very specific circumstances.
9. The model of S-subunit of EcoR124I has allowed to predict a number of functionally relevant residues potentially indirectly involved in DNA binding.

10. Point mutations in the conserved domain of HsdS in positions K184 and K384 make the helices more rigid, probably trapping the HsdS subunit in a conformation in which the MS-complex cannot assemble properly with the HsdRsubunit, resulting in a loss of restriction activity.

Future Work

A full description of the translocation cycle by enhanced sampling methods would be a logical next step. Recently our group was able to crystallize the missing 150 residue of HsdR C-terminal and the structure has been solved. Our preliminary results, that go beyond the scope of this thesis, show an interesting motion of the C-terminal with respect to the rest of the protein and suggest a path of for DNA binding and translocation over the enzyme as well. Assembly of the full HsdR subunit with the metyltransferase is the next aim, unfortunately requiring a high-resolution structure of HsdM subunit, but once achieved it would allow us to comprehend the *in vitro* interactions among the subunits. The crystallographic studies of the HsdM subunit is on-going and the in-house X-ray diffraction analysis shows that initial HsdM crystals diffract although the crystals still need to be optimized before the full dataset can be collected at the synchrotron. This could be a very promising experimental result that might lead us to carry out MD simulations of the HsdM crystal structure with the computational model of HsdS subunit. On the other hand structure of full pentameric complex continues to be a significant challenge and is being studied by cryo-electron microscopy experiments.

5.Publications

5.1

J Mol Model (2014) 20:2334
DOI 10.1007/s00894-014-2334-1

ORIGINAL PAPER

Interdomain communication in the endonuclease/motor subunit of type I restriction-modification enzyme EcoR124I

Dhiraj Sinha · Katsiaryna Shamayeva · Vyas Ramasubramani · David Řeha · Vitali Bialevich · Morteza Khabiri · Alena Guzanová · Niv Milbar · Marie Weiserová · Eva Csefalvay · Jannette Carey · Rüdiger Ettrich

Received: 19 December 2013 / Accepted: 3 June 2014 / Published online: 28 June 2014

Abstract Restriction-modification systems protect bacteria from foreign DNA. Type I restriction-modification enzymes are multifunctional heteromeric complexes with DNA-cleavage and ATP-dependent DNA translocation activities located on endonuclease/motor subunit HsdR. The recent structure of the first intact motor subunit of the type I restriction enzyme from plasmid EcoR124I suggested a mechanism by which stalled translocation triggers DNA cleavage via a lysine residue on the endonuclease domain that contacts ATP bound between the two helicase domains. In the present work, molecular dynamics simulations are used to explore this proposal. Molecular dynamics simulations suggest that the Lys–ATP contact alternates with a contact with a nearby loop housing the conserved QxxxY motif that had been implicated in DNA cleavage. This model is tested here using in vivo and in vitro experiments. The results indicate how local interactions are transduced to domain motions within the endonuclease/motor subunit.

Keywords DNA restriction enzymes · Molecular modeling · QM/MM calculations · Principal components analysis · *E. coli* · Multisubunit enzyme complex · Correlated loop motions

Dhiraj Sinha and Katsiaryna Shamayeva contributed equally.

This paper belongs to Topical Collection MIB 2013 (Modeling Interactions in Biomolecules VI)

Electronic supplementary material The online version of this article (doi:10.1007/s00894-014-2334-1) contains supplementary material, which is available to authorized users.

D. Sinha · K. Shamayeva · D. Řeha · V. Bialevich · M. Khabiri · E. Csefalvay · J. Carey · R. Ettrich
Institute of Nanobiology and Structural Biology, Global Change Research Center, Academy of Sciences of the Czech Republic, Zamek 136, 373 33 Nove Hrady, Czech Republic

D. Sinha · K. Shamayeva · D. Řeha · V. Bialevich · R. Ettrich (✉)
Faculty of Sciences, University of South Bohemia in Ceske Budejovice, Zamek 136, 373 33 Nove Hrady, Czech Republic
e-mail: ettrich@nh.cas.cz

V. Ramasubramani · N. Milbar · J. Carey
Chemistry Department, Princeton University, Princeton, NJ 08544-1009, USA

A. Guzanová · M. Weiserová
Laboratory of Molecular Genetics of Bacteria, Institute of Microbiology, Academy of Sciences of the Czech Republic, Videňská 1083, 142 20 Praha 4, Czech Republic

Introduction

Restriction-modification systems protect bacteria from foreign DNA. Type II restriction enzymes, which are relatively simple, are widely used in molecular biology and are well understood. Type I enzymes, which were discovered first, are not used commercially because of their complicated subunit structure and cleavage behavior. They are encoded by the host specificity of DNA (*hsd*) genes; for a review see [1]. Two copies of the HsdM methylation subunit and one copy of specificity subunit HsdS bind at a specific asymmetric DNA site and form a methyltransferase complex that modifies the host DNA, protecting it from cleavage. When invading DNA is detected by the methyltransferase based on its methylation status, two HsdR endonuclease/motor subunits are recruited to the complex, which then translocates up to thousands of duplex base pairs through the stationary enzyme without unwinding the DNA, consuming ~1 ATP per base pair. If translocation is blocked, the endonuclease activity of HsdR is triggered, and DNA is cleaved nonspecifically at sites distant from the original binding site. These unusual activities have prompted many years of genetic and biochemical study aimed at understanding the molecular

5.2

J Mol Model (2014) 20:2400
DOI 10.1007/s00894-014-2400-8

ORIGINAL PAPER

Molecular dynamics comparison of *E. coli* WrbA apoprotein and holoprotein

David Reha · Balasubramanian Harish · Dhiraj Sinha · Zdenek Kukacka · James McSally · Olga Ettrichova · Petr Novak · Jannette Carey · Rüdiger Ettrich

Received: 29 December 2013 / Accepted: 23 July 2014 / Published online: 26 August 2014

Abstract WrbA is a novel multimeric flavodoxin-like protein of unknown function. A recent high-resolution X-ray crystal structure of *E. coli* WrbA holoprotein revealed a methionine sulfoxide residue with full occupancy in the FMN-binding site, a finding that was confirmed by mass spectrometry. In an effort to evaluate whether methionine sulfoxide may have a role in WrbA function, the present analyses were undertaken using molecular dynamics simulations in combination with further mass spectrometry of the protein. Methionine sulfoxide formation upon reconstitution of purified apoWrbA with oxidized FMN is fast as judged by kinetic mass spectrometry, being complete in ~5 h and resulting in complete conversion at the active-site methionine with minor extents of conversion at heterogeneous second sites. Analysis of methionine oxidation states during purification of holoWrbA from bacterial cells reveals that methionine is not oxidized prior to reconstitution, indicating that methionine sulfoxide is unlikely to be relevant to the function of WrbA in vivo. Although the simulation results, the first reported for WrbA, led to no hypotheses about the role of methionine sulfoxide that could be tested experimentally, they elucidated the origins of the two major differences between apo- and holoWrbA crystal structures, an alteration of inter-subunit distance and a rotational shift within the tetrameric assembly.

Keywords Global motions · Force field parametrization · Binding site volume · Electrostatic potential surface · NAD(P)H:quinone oxidoreductase

This paper belongs to Topical Collection MIB 2013 (Modeling Interactions in Biomolecules VI)

Electronic supplementary material The online version of this article (doi:10.1007/s00894-014-2400-8) contains supplementary material, which is available to authorized users.

D. Reha (✉) · D. Sinha · O. Ettrichova · R. Ettrich
Institute of Nanobiology and Structural Biology, Global Change Research Center, Academy of Sciences of the Czech Republic, Zamek 136, 373 33 Nove Hrady, Czech Republic
e-mail: reha@nh.cas.cz

D. Reha · D. Sinha · R. Ettrich
Faculty of Sciences, University of South Bohemia in Ceske Budejovice, Zamek 136, 373 33 Nove Hrady, Czech Republic

B. Harish · J. McSally · J. Carey
Chemistry Department, Princeton University, Princeton, NJ 08544-1009, USA

Z. Kukacka · P. Novak
Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Praha 4, Czech Republic

Z. Kukacka · P. Novak
Faculty of Science, Charles University in Prague, Albertov 6, 128 43 Praha 2, Czech Republic

Introduction

E. coli protein WrbA ([1]; Fig. 1) is the founding member of a class of novel flavoproteins conserved from bacteria to higher plants [2] whose exact functional role is still unknown [3]. Like the flavodoxins to which it is distantly related, WrbA can use its physiological cofactor FMN [4] to transfer electrons [5], but unlike the flavodoxins WrbA transfers two electrons at a time [6] from NADH to quinones. This function is similar to that of the mammalian FAD-dependent NAD (P) H:quinone oxidoreductases [7, 8] to which WrbA is structurally related, suggesting a role in quinone detoxification and oxidative-stress defense. Unlike both these related proteins, WrbA in solution participates in a dimer-tetramer equilibrium [4]. It has thus been suggested [9, 10] that WrbA represents a structural and functional bridge between the monomeric FMN-dependent bacterial flavodoxins and the dimeric FAD-dependent mammalian NAD (P) H-quinone oxidoreductases.

5.3

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Quantum Calculations Indicate Effective Electron Transfer between FMN and Benzoquinone in a New Crystal Structure of *Escherichia coli* WrbA

Oksana Degtjarik,[†,‡,#] Jiří Brynda,[§] Olga Ettrichova,[†] Michal Kuty,[†,‡] Dhiraj Sinha,[†,‡] Ivana Kuta Smatanova,[†,‡] Jannette Carey,[†,∥] Rüdiger Ettrich,[†,‡] and David Řeha*[,†,‡]

[†]Center for Nanobiology and Structural Biology, Institute of Microbiology, Academy of Sciences of the Czech Republic, Zamek 136, CZ-373 33 Nove Hrady, Czech Republic

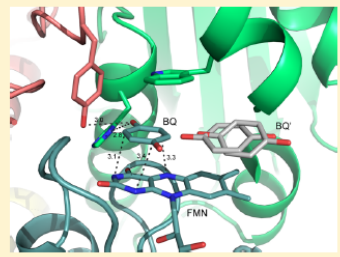
[‡]Faculty of Sciences, University of South Bohemia in Ceske Budejovice, Zamek 136, CZ-373 33 Nove Hrady, Czech Republic

[§]Institute of Molecular Genetics and Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 16610 Prague 6, Czech Republic

[∥]Chemistry Department, Princeton University, Princeton, New Jersey 08544-1009, United States

Supporting Information

ABSTRACT: Quantum mechanical calculations using the Marcus equation are applied to compare the electron-transfer probability for two distinct crystal structures of the *Escherichia coli* protein WrbA, an FMN-dependent NAD(P)H:quinone oxidoreductase, with the bound substrate benzoquinone. The calculations indicate that the position of benzoquinone in a new structure reported here and solved at 1.33 Å resolution is more likely to be relevant for the physiological reaction of WrbA than a previously reported crystal structure in which benzoquinone is shifted by ∼5 Å. Because the true electron-acceptor substrate for WrbA is not yet known, the present results can serve to constrain computational docking attempts with potential substrates that may aid in identifying the natural substrate(s) and physiological role(s) of this enzyme. The approach used here highlights a role for quantum mechanical calculations in the interpretation of protein crystal structures.



INTRODUCTION

The FMN-dependent NAD(P)H:quinone oxidoreductase WrbA isolated from *Escherichia coli* (Figure 1) is the founding member of an unusual family of flavodoxin-like proteins.[1] WrbA-like proteins combine features found both in monomeric bacterial flavodoxins that use FMN as cofactor and in dimeric eukaryotic FAD-dependent oxidoreductases (Nqos).[2] Like the flavodoxins, WrbAs adopt the canonical α/β twisted open-sheet fold and use FMN as physiological cofactor,[1] but like the Nqos they have a characteristic short sequence insertion after β2 that forms an additional beta-alpha secondary structural element lying outside the canonical fold.[2−6] Like the Nqos, WrbAs populate a dimeric state in solution, but they also undergo a facile dynamic equilibrium to form active tetramers.[2] Unexpectedly, the assembly of WrbA multimers relies not on their characteristic sequence insertion relative to the flavodoxins, but on secondary structural elements in common with the flavodoxin fold that in WrbA appear to be adapted for subunit interaction.[5] The insertion instead is located at the two polar regions of the tetrahedral assembly, where it contributes to forming a hydrophobic channel leading from the protein surface to the active sites.[2]

Each cavernous active site is formed by residues from three WrbA subunits,[2] indicating that tetramers are the active species. The FMN isoalloxazine ring is stacked between aromatic residues similarly as in flavodoxins, although the topological origin of the active-site residues differs.[5] Biochemical assays demonstrate reducing activity of WrbA toward short- and long-chain quinones[7,8] including benzoquinone (BQ) that is used in a standard assay with NADH as electron donor. The ping-pong kinetic mechanisms of *E. coli* WrbA[8] and the Nqos,[9,10] together with computational docking experiments,[8] indicate that the electron donor substrate NADH and the electron acceptor substrate BQ do not occupy the active site simultaneously. The hydrophobic channel connecting the active site to the protein surface has been proposed as a possible conduit for long-chain, membrane-bound quinones.[2] However, the natural substrate(s) and physiological role(s) of WrbA are unknown, and its ubiquity among bacteria and plants remains unexplained.

One possible route to understanding the natural role of WrbA is to evaluate the suitability of potential physiological

Received: December 7, 2015
Revised: May 15, 2016
Published: May 16, 2016

DOI: 10.1021/acs.jpcb.5b11958
J. Phys. Chem. B 2016, 120, 4867−4877

5.4

The helical domain of the motor subunit of EcoR124I participates in ATPase activity and dsDNA translocation

Bialevich V[1,2,#], Sinha D[1,2,#], Shamayeva K[1,2], Guzanova A[3], Csefalvay E[1], Carey J[1,4], Weiserova M[3], Ettrich R[1,2]*

5.5

The role of motif III and its extended region in positioning the two helicase domains in the motor subunit of the restriction–modification system EcoR124I

Dhiraj Sinha[1,2,#], VitaliBialevich[1,2,#], Katsiaryna Shamayeva[1,2], Alena Guzanova[3], Alexandra Sisakova[4], Eva Csfalvay[1], Lumir Krejci[4,5,6], Jannette Carey[1,7], Marie Weiserova[3], Rüdiger Ettrich[1,2,]*

5.6

Involvement of the conservative domains of the DNA recognition subunit HsdS in assembly of function EcoR124I endonuclease

Guzanova A, Sinha D, Pal K Sudhir, Ettrich R and Weiserova M

6.References

Ahmad, I. and D. N. Rao (1996). "Functional analysis of conserved motifs in EcoP15I DNA methyltransferase." J Mol Biol **259**(2):229-240.

Bannister, D. and S. W. Glover (1968). "Restriction and modification of bacteriophages by R+ strains of Escherichia coli K12." Biochem Biophys Res Commun **30**(6):735-738.

Barcus, V. A., A. J. Titheradge and N. E. Murray (1995). "The diversity of alleles at the hsd locus in natural populations of Escherichia coli." Genetics **140**(4):1187-1197.

Benz, J., H. Trachsel and U. Baumann (1999). "Crystal structure of the ATPase domain of translation initiation factor 4A from Saccharomyces cerevisiae--the prototype of the DEAD box protein family." Structure **7**(6):671-679.

Bertani, G. and J. J. Weigle (1953). "Host controlled variation in bacterial viruses." J Bacteriol **65**(2):113-121.

Bhaya, D., M. Davison and R. Barrangou (2011). "CRISPR-Cas systems in bacteria and archaea:versatile small RNAs for adaptive defense and regulation." Annu Rev Genet **45**:273-297.

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

Bickle, T. A. and D. H. Kruger (1993). "Biology of DNA restriction." Microbiol Rev **57**(2):434-450.

Blower, T. R., X. Y. Pei, F. L. Short, P. C. Fineran, D. P. Humphreys, B. F. Luisi and G. P. Salmond (2011). "A processed noncoding RNA regulates an altruistic bacterial antiviral system." Nat Struct Mol Biol **18**(2):185-190.

Blum, S., S. R. Schmid, A. Pause, P. Buser, P. Linder, N. Sonenberg and H. Trachsel (1992). "ATP hydrolysis by initiation factor 4A is required for translation initiation in Saccharomyces cerevisiae." Proc Natl Acad Sci USA **89**(16):7664-7668.

Boyer, H. W. (1971). "DNA restriction and modification mechanisms in bacteria." Annu Rev Microbiol **25**:153-176.

Brammar, W. J., N. E. Murray and S. Winton (1974). "Restriction of lambda trp bacteriophages by Escherichia coli K." J Mol Biol **90**(4):633-647.

Bruccoleri, R. E. and M. Karplus (1987). "Prediction of the folding of short polypeptide segments by uniform conformational sampling." Biopolymers **26**(1):137-168.

Bujnicki, J. M. and L. Rychlewski (2001). "Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles." J Mol Microbiol Biotechnol **3**(1):69-72.

Calisto, B. M., O. Q. Pich, J. Pinol, I. Fita, E. Querol and X. Carpena (2005). "Crystal structure of a putative type I restriction-modification S subunit from Mycoplasma genitalium." J Mol Biol **351**(4):749-762.

Callow, P., A. Sukhodub, J. E. Taylor and G. G. Kneale (2007). "Shape and subunit organisation of the DNA methyltransferase M.AhdI by small-angle neutron scattering." J Mol Biol **369**(1):177-185.

Caruthers, J. M., E. R. Johnson and D. B. McKay (2000). "Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase." Proc Natl Acad Sci USA **97**(24):13080-13085.

Caruthers, J. M. and D. B. McKay (2002). "Helicase structure and mechanism." Curr Opin Struct Biol **12**(1):123-133.

Chang, S. and S. N. Cohen (1977). "In vivo site-specific genetic recombination promoted by the EcoRI restriction endonuclease." Proc Natl Acad Sci USA **74**(11):4811-4815.

Chin, V., V. Valinluck, S. Magaki and J. Ryu (2004). "KpnBI is the prototype of a new family (IE) of bacterial type I restriction-modification system." Nucleic Acids Res **32**(18):e138.

Chothia, C., A. M. Lesk, M. Levitt, A. G. Amit, R. A. Mariuzza, S. E. Phillips and R. J. Poljak (1986). "The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure." Science **233**(4765):755-758.

Cordin, O., N. K. Tanner, M. Doere, P. Linder and J. Banroques (2004). "The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity." EMBO J **23**(13):2478-2487.

Csefalvay, E., M. Lapkouski, A. Guzanova, L. Csefalvay, T. Baikova, I. Shevelev, V. Bialevich, K. Shamayeva, P. Janscak, I. Kuta Smatanova, S. Panjikar, J. Carey, M. Weiserova and R. Ettrich (2015). "Functional coupling of duplex translocation to DNA cleavage in a type I restriction enzyme." PLoS One **10**(6):e0128700.

Danna, K. and D. Nathans (1971). "Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae." Proc Natl Acad Sci USA **68**(12):2913-2917.

Davies, G. P., P. Kemp, I. J. Molineux and N. E. Murray (1999). "The DNA translocation and ATPase activities of restriction-deficient mutants of Eco KI." J Mol Biol **292**(4):787-796.

Davies, G. P., I. Martin, S. S. Sturrock, A. Cronshaw, N. E. Murray and D. T. Dryden (1999). "On the structure and operation of type I DNA restriction enzymes." J Mol Biol **290**(2):565-579.

Davies, G. P., L. M. Powell, J. L. Webb, L. P. Cooper and N. E. Murray (1998). "EcoKI with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities." Nucleic Acids Res **26**(21):4828-4836.

Deane, C. M. and T. L. Blundell (2000). "A novel exhaustive search algorithm for predicting the conformation of polypeptide segments in proteins." Proteins**40**(1):135-144.

Delagoutte, E. and P. H. von Hippel (2002). "Helicase mechanisms and the coupling of helicases within macromolecular machines. Part I:Structures and properties of isolated helicases." Q Rev Biophys **35**(4):431-478.

Donmez, I. and S. S. Patel (2006). "Mechanisms of a ring shaped helicase." Nucleic Acids Res **34**(15):4216-4224.

Dreier, J. and T. A. Bickle (1996). "ATPase activity of the type IC restriction-modification system EcoR124II." J Mol Biol **257**(5):960-969.

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

Dryden, D. T., L. P. Cooper and N. E. Murray (1993). "Purification and characterization of the methyltransferase from the type 1 restriction and modification system of Escherichia coli K12." J Biol Chem **268**(18):13228-13236.

Dryden, D. T., L. P. Cooper, P. H. Thorpe and O. Byron (1997). "The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications." Biochemistry **36**(5):1065-1076.

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

Dryden, D. T., S. S. Sturrock and M. Winter (1995). "Structural modelling of a type I DNA methyltransferase." Nat Struct Biol **2**(8):632-635.

Dürr, H., C. Korner, M. Muller, V. Hickmann and K. P. Hopfner (2005). "X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA." Cell **121**(3):363-373.

Dussoix, D. and W. Arber (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.

Endlich, B. and S. Linn (1985). "The DNA restriction endonuclease of Escherichia coli B. I. Studies of the DNA translocation and the ATPase activities." J Biol Chem **260**(9):5720-5728.

Endlich, B. and S. Linn (1985). "The DNA restriction endonuclease of Escherichia coli B. II. Further studies of the structure of DNA intermediates and products." J Biol Chem **260**(9):5729-5738.

Eskin, B. and S. Linn (1972). "The deoxyribonucleic acid modification and restriction enzymes of Escherichia coli B. II. Purification, subunit structure, and catalytic properties of the restriction endonuclease." J Biol Chem **247**(19):6183-6191.

Fine, R. M., H. Wang, P. S. Shenkin, D. L. Yarmush and C. Levinthal (1986). "Predicting antibody hypervariable loop conformations. II:Minimization and molecular dynamics studies of MCPC603 from many randomly generated loop conformations." Proteins **1**(4):342-362.

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

Firman, K. and M. D. Szczelkun (2000). "Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement." EMBO J **19**(9):2094-2102.

Fiser, A., R. K. Do and A. Sali (2000). "Modeling of loops in protein structures." Protein Sci **9**(9):1753-1773.

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

Glover, S. W. (1970). "Functional analysis of host-specificity mutants in Escherichia coli." Genet Res **15**(2):237-250.

Glover, S. W. and C. Colson (1969). "Genetics of host-controlled restriction and modification in Escherichia coli." Genet Res **13**(2):227-240.

Gomez, P. and A. Buckling (2011). "Bacteria-phage antagonistic coevolution in soil." Science **332**(6025):106-109.

Gorbalenya, A. E. and E. V. Koonin (1991). "Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain." FEBS Lett **291**(2):277-281.

Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko and V. M. Blinov (1988). "A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination." FEBS Lett **235**(1-2):16-24.

Greer, J. (1980). "Model for haptoglobin heavy chain based upon structural homology." Proc Natl Acad Sci USA **77**(6):3393-3397.

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

Haberman, A. (1974). "The bacteriophage P1 restriction endonuclease." J Mol Biol**89**(4):545-563.

Hall, M. C. and S. W. Matson (1999). "Helicase motifs:the engine that powers DNA unwinding." Mol Microbiol **34**(5):867-877.

Hattman, S., J. Wilkinson, D. Swinton, S. Schlagman, P. M. Macdonald and G. Mosig (1985). "Common evolutionary origin of the phage T4 dam and host Escherichia coli dam DNA-adenine methyltransferase genes." J Bacteriol **164**(2):932-937.

Horiuchi, K. and N. D. Zinder (1972). "Cleavage of bacteriophage fl DNA by the restriction enzyme of Escherichia coli B." Proc Natl Acad Sci USA **69**(11):3220-3224.

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

Hughes, S. G. (1977). "A map of the cleavage sites for endonuclease AvaI in the chromosome of bacteriophage lambda." Biochem J **163**(3):503-509.

Ilyina, T. V., A. E. Gorbalenya and E. V. Koonin (1992). "Organization and evolution of bacterial and bacteriophage primase-helicase systems." J Mol Evol **34**(4):351-357.

Hyman, P. and Abedon, S.T. (2010). Bacteriophage host range and bacterial resistance. Adv. Appl. Microbiol. **70**:217–48.

Irizarry, R. A., C. Ladd-Acosta, B. Carvalho, H. Wu, S. A. Brandenburg, J. A. Jeddeloh, B. Wen and A. P. Feinberg (2008). "Comprehensive high-throughput arrays for relative methylation (CHARM)." Genome Res **18**(5):780-790.

Janscak, P. and T. A. Bickle (1998). "The DNA recognition subunit of the type IB restriction-modification enzyme EcoAI tolerates circular permutions of its polypeptide chain." J Mol Biol **284**(4):937-948.

Janscak, P. and T. A. Bickle (2000). "DNA supercoiling during ATP-dependent DNA translocation by the type I restriction enzyme EcoAI." J Mol Biol**295**(4):1089-1099.

Janscak, P., D. T. Dryden and K. Firman (1998). "Analysis of the subunit assembly of the typeIC restriction-modification enzyme EcoR124I." Nucleic Acids Res **26**(19):4439-4445.

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

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

Janscak, P., M. Weiserova, J. Hubacek, I. Holubova, C. F. Dutta and K. Firman (2000). "Two temperature-sensitive mutations in the DNA binding subunit of EcoKI with differing properties." FEMS Microbiol Lett **182**(1):99-104.

Janulaitis, A., R. Vaisvila, A. Timinskas, S. Klimasauskas and V. Butkus (1992). "Cloning and sequence analysis of the genes coding for Eco57I type IV restriction-modification enzymes." Nucleic Acids Res **20**(22):6051-6056.

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

Johnson, B. B., K. S. Dahl, I. Tinoco, Jr., V. I. Ivanov and V. B. Zhurkin (1981). "Correlations between deoxyribonucleic acid structural parameters and calculated circular dichroism spectra." Biochemistry **20**(1):73-78.

Kauc, L. (1992). "Determination of genome size of Haemophilus influenzae Sb:analysis of DNA restriction fragments." Acta Microbiol Pol **41**(1-2):13-24.

Kauc, L. and A. Piekarowicz (1978). "Purification and properties of a new restriction endonuclease from Haemophilus influenzae Rf." Eur J Biochem **92**(2):417-426.

Kelly, T. J., Jr. and H. O. Smith (1970). "A restriction enzyme from Hemophilus influenzae. II." J Mol Biol **51**(2):393-409.

Kennaway, C. K., A. Obarska-Kosinska, J. H. White, I. Tuszynska, L. P. Cooper, J. M. Bujnicki, J. Trinick and D. T. Dryden (2009). "The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein." Nucleic Acids Res **37**(3):762-770.

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

Kessler, C. and V. Manta (1990). "Specificity of restriction endonucleases and DNA modification methyltransferases a review (Edition 3)." Gene **92**(1-2):1-248.

Kim, J. S., A. DeGiovanni, J. Jancarik, P. D. Adams, H. Yokota, R. Kim and S. H. Kim (2005). "Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications." Proc Natl Acad Sci USA **102**(9):3248-3253.

Kneale, G. G. (1994). "A symmetrical model for the domain structure of type I DNA methyltransferases." J Mol Biol **243**(1):1-5.

Kobayashi, I. (2001). "Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution." Nucleic Acids Res **29**(18):3742-3756.

Kosinski, J., M. J. Gajda, I. A. Cymerman, M. A. Kurowski, M. Pawlowski, M. Boniecki, A. Obarska, G. Papaj, P. Sroczynska-Obuchowicz, K. L. Tkaczuk, P. Sniezynska, J. M. Sasin, A. Augustyn, J. M. Bujnicki and M. Feder (2005). "FRankenstein becomes a cyborg:the automatic recombination and realignment of fold recognition models in CASP6." Proteins **61** Suppl **7**:106-113.

Kuhnlein, U. and W. Arber (1972). "Host specificity of DNA produced by Escherichia coli. XV. The role of nucleotide methylation in in vitro B-specific modification." J Mol Biol **63**(1):9-19.

Kusano, K., T. Naito, N. Handa and I. Kobayashi (1995). "Restriction-modification systems as genomic parasites in competition for specific sequences." Proc Natl Acad Sci USA **92**(24):11095-11099.

Labrie, S. J., J. E. Samson and S. Moineau (2010). "Bacteriophage resistance mechanisms." Nat Rev Microbiol **8**(5):317-327.

Lapkouski, M., S. Panjikar, P. Janscak, I. K. Smatanova, J. Carey, R. Ettrich and E. Csefalvay (2009). "Structure of the motor subunit of type I restriction-modification complex EcoR124I." Nat Struct Mol Biol **16**(1):94-95.

Lewis, R., H. Dürr, K. P. Hopfner and J. Michaelis (2008). "Conformational changes of a Swi2/Snf2 ATPase during its mechano-chemical cycle." Nucleic Acids Res **36**(6):1881-1890.

Linn, S. and W. Arber (1968). "Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form." Proc Natl Acad Sci USA **59**(4):1300-1306.

Lohman, T. M. and K. P. Bjornson (1996). "Mechanisms of helicase-catalyzed DNA unwinding." Annu Rev Biochem **65**:169-214.

Luria, S. E. (1953). "Host-induced modifications of viruses." Cold Spring Harb Symp Quant Biol **18**:237-244.

MacWilliams, M. P. and T. A. Bickle (1996). "Generation of new DNA binding specificity by truncation of the type IC EcoDXXI hsdS gene." EMBO J **15**(17):4775-4783.

Matson, S. W., D. W. Bean and J. W. George (1994). "DNA helicases:enzymes with essential roles in all aspects of DNA metabolism." Bioessays**16**(1):13-22.

McCammon, J. A. and M. Karplus (1977). "Internal motions of antibody molecules." Nature **268**(5622):765-766.

McClelland, S. E., D. T. Dryden and M. D. Szczelkun (2005). "Continuous assays for DNA translocation using fluorescent triplex dissociation:application to type I restriction endonucleases." J Mol Biol **348**(4):895-915.

Mernagh, D. R., P. Janscak, K. Firman and G. G. Kneale (1998). "Protein-protein and protein-DNA interactions in the type I restriction endonuclease R.EcoR124I." Biol Chem **379**(4-5):497-503.

Mernagh, D. R., I. A. Taylor and G. G. Kneale (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.

Meynell, E. and N. Datta (1966). "The relation of resistance transfer factors to the F-factor (sex-factor) of Escherichia coli K12." Genet Res **7**(1):134-140.

Moult, J. and M. N. James (1986). "An algorithm for determining the conformation of polypeptide segments in proteins by systematic search." Proteins **1**(2):146-163.

Murray, N. E. (2000). "Type I restriction systems:sophisticated molecular machines (a legacy of Bertani and Weigle)." Microbiol Mol Biol Rev **64**(2):412-434.

Murray, N. E., P. L. Batten and K. Murray (1973). "Restriction of bacteriophage lambda by Escherichia coli K." J Mol Biol **81**(3):395-407.

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

Murray, N. E., J. A. Gough, B. Suri and T. A. Bickle (1982). "Structural homologies among type I restriction-modification systems." EMBO J **1**(5):535-539.

Nagaraja, V., J. C. Shepherd and T. A. Bickle (1985). "A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity." Nature **316**(6026):371-372.

Nagaraja, V., J. C. Shepherd, T. Pripfl and T. A. Bickle (1985). "Two type I restriction enzymes from Salmonella species. Purification and DNA recognition sequences." J Mol Biol **182**(4):579-587.

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

Obarska, A., A. Blundell, M. Feder, S. Vejsadova, E. Sisakova, M. Weiserova, J. M. Bujnicki and K. Firman (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.

Ordway, J. M., J. A. Bedell, R. W. Citek, A. Nunberg, A. Garrido, R. Kendall, J. R. Stevens, D. Cao, R. W. Doerge, Y. Korshunova, H. Holemon, J. D. McPherson, N. Lakey, J. Leon, R. A. Martienssen and J. A. Jeddeloh (2006). "Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets." Carcinogenesis **27**(12):2409-2423.

Ostrander, E. A., P. Benedetti and J. C. Wang (1990). "Template supercoiling by a chimera of yeast GAL4 protein and phage T7 RNA polymerase." Science **249**(4974):1261-1265.

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

Powell, L. M., D. T. Dryden and N. E. Murray (1998). "Sequence-specific DNA binding by EcoKI, a type IA DNA restriction enzyme." J Mol Biol **283**(5):963-976.

Powell, L. M., D. T. Dryden, D. F. Willcock, R. H. Pain and N. E. Murray (1993). "DNA recognition by the EcoK methyltransferase. The influence of DNA methylation and the cofactor S-adenosyl-L-methionine." J Mol Biol **234**(1):60-71.

Powell, L. M. and N. E. Murray (1995). "S-adenosyl methionine alters the DNA contacts of the EcoKI methyltransferase." Nucleic Acids Res **23**(6):967-974.

Price, C., J. Lingner, T. A. Bickle, K. Firman and S. W. Glover (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.

Price, C., T. Pripfl and T. A. Bickle (1987). "EcoR124 and EcoR124/3:the first members of a new family of type I restriction and modification systems." Eur J Biochem **167**(1):111-115.

Price, C., J. C. Shepherd and T. A. Bickle (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.

Rao, D. N., S. Saha and V. Krishnamurthy (2000). "ATP-dependent restriction enzymes." Prog Nucleic Acid Res Mol Biol **64**:1-63.

Redaschi, N. and T. A. Bickle (1996). "Posttranscriptional regulation of EcoP1I and EcoP15I restriction activity." J Mol Bio **257**(4):790-803.

Roberts, R. J., M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. Degtyarev, D. T. Dryden, K. Dybvig, K. Firman, E. S. Gromova, R. I. Gumport, S. E. Halford, S. Hattman, J. Heitman, D. P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T. R. Klaenhammer, I. Kobayashi, H. Kong, D. H. Kruger, S. Lacks, M. G. Marinus, M. Miyahara, R. D. Morgan, N. E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D. N. Rao, N. Reich, V. E. Repin, E. U. Selker, P. C. Shaw, D. C. Stein, B. L. Stoddard, W. Szybalski, T. A. Trautner, J. L. Van Etten, J. M. Vitor, G. G. Wilson and S. Y. Xu (2003). "A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes." Nucleic Acids Res **31**(7):1805-1812.

Roberts, R. J., T. Vincze, J. Posfai and D. Macelis (2015). "REBASE--a database for DNA restriction and modification:enzymes, genes and genomes." Nucleic Acids Res **43**(Database issue):D298-299.

Rocak, S. and P. Linder (2004). "DEAD-box proteins:the driving forces behind RNA metabolism." Nat Rev Mol Cell Biol **5**(3):232-241.

Rogers, G. W., Jr., A. A. Komar and W. C. Merrick (2002). "eIF4A:the godfather of the DEAD box helicases." Prog Nucleic Acid Res Mol Biol **72**:307-331.

Rohozinski, J., L. E. Girton and J. L. Van Etten (1989). "Chlorella viruses contain linear nonpermuted double-stranded DNA genomes with covalently closed hairpin ends." Virology **168**(2):363-369.

Roossinck, M. J. (2011). "The good viruses:viral mutualistic symbioses." Nat Rev Microbiol **9**(2):99-108.

Rosamond, J., B. Endlich and S. Linn (1979). "Electron microscopic studies of the mechanism of action of the restriction endonuclease of Escherichia coli B." J Mol Biol **129**(4):619-635.

Rozen, F., J. Pelletier, H. Trachsel and N. Sonenberg (1989). "A lysine substitution in the ATP-binding site of eucaryotic initiation factor 4A abrogates nucleotide-binding activity." Mol Cell Biol **9**(9):4061-4063.

Schluckebier, G., J. Labahn, J. Granzin and W. Saenger (1998). "M.TaqI:possible catalysis via cation-pi interactions in N-specific DNA methyltransferases." Biol Chem**379**(4-5):389-400.

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

Seidel, R., J. van Noort, C. van der Scheer, J. G. Bloom, N. H. Dekker, C. F. Dutta, A. Blundell, T. Robinson, K. Firman and C. Dekker (2004). "Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I." Nat Struct Mol Biol **11**(9):838-843.

Sharp, P. A., B. Sugden and J. Sambrook (1973). "Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose--ethidium bromide electrophoresis." Biochemistry **12**(16):3055-3063.

Simons, K. T., R. Bonneau, I. Ruczinski and D. Baker (1999). "Ab initio protein structure prediction of CASP III targets using ROSETTA." Proteins Suppl **3**:171-176.

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

Sisakova, E., L. K. Stanley, M. Weiserova and M. D. Szczelkun (2008). "A RecB-family nuclease motif in the Type I restriction endonuclease EcoR124I." Nucleic Acids Res **36**(12):3939-3949.

Smith, H. O. and D. Nathans (1973). "Letter:A suggested nomenclature for bacterial host modification and restriction systems and their enzymes." J Mol Biol **81**(3):419-423.

Smith, H. O. and K. W. Wilcox (1970). "A restriction enzyme from Hemophilus influenzae. I. Purification and general properties." J Mol Biol**51**(2):379-391.

Soultanas, P. and D. B. Wigley (2001). "Unwinding the 'Gordian knot' of helicase action." Trends Biochem Sci **26**(1):47-54.

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

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

Story, R. M., H. Li and J. N. Abelson (2001). "Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii." Proc Natl Acad Sci USA **98**(4):1465-1470.

Story, R. M. and T. A. Steitz (1992). "Structure of the recA protein-ADP complex." Nature **355**(6358):374-376.

Strick, T. R., J. F. Allemand, D. Bensimon and V. Croquette (1998). "Behavior of supercoiled DNA." Biophys J **74**(4):2016-2028.

Studier, F. W. and P. K. Bandyopadhyay (1988). "Model for how type I restriction enzymes select cleavage sites in DNA." Proc Natl Acad Sci USA **85**(13):4677-4681.

Sturrock, S. S. and D. T. Dryden (1997). "A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes." Nucleic Acids Res **25**(17):3408-3414.

Su, T. J., M. R. Tock, S. U. Egelhaaf, W. C. Poon and D. T. Dryden (2005). "DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping." Nucleic Acids Res **33**(10):3235-3244.

Summers, N. L. and M. Karplus (1990). "Modeling of globular proteins. A distance-based data search procedure for the construction of insertion/deletion regions and Pro----non-Pro mutations." J Mol Biol **216**(4):991-1016.

Suri, B. and T. A. Bickle (1985). "EcoA:the first member of a new family of type I restriction modification systems. Gene organization and enzymatic activities." J Mol Biol **86**(1):77-85.

Szczelkun, M. D. (2000). "How to proteins move along DNA? Lessons from type-I and type-III restriction endonucleases." Essays Biochem **35**:131-143.

Szczelkun, M. D. (2002). "Kinetic models of translocation, head-on collision, and DNA cleavage by type I restriction endonucleases." Biochemistry **41**(6):2067-2074.

Szczelkun, M. D., M. S. Dillingham, P. Janscak, K. Firman and S. E. Halford (1996). "Repercussions of DNA tracking by the type IC restriction endonuclease

EcoR124I on linear, circular and catenated substrates." EMBO J **15**(22):6335-6347.

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

Tanner, N. K., O. Cordin, J. Banroques, M. Doere and P. Linder (2003). "The Q motif:a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis." Mol Cell **11**(1):127-138.

Taylor, I., J. Patel, K. Firman and G. Kneale (1992). "Purification and biochemical characterisation of the EcoR124 type I modification methylase." Nucleic Acids Res **20**(2):179-186.

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

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

Taylor, J. E., P. Callow, A. Swiderska and G. G. Kneale (2010). "Structural and functional analysis of the engineered type I DNA methyltransferase EcoR124I(NT)." J Mol Biol **398**(3):391-399.

Titheradge, A. J., D. Ternent and N. E. Murray (1996). "A third family of allelic hsd genes in Salmonella enterica:sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA." Mol Microbiol **22**(3):437-447.

van der Spoel, D. and H. J. Berendsen (1997). "Molecular dynamics simulations of Leu-enkephalin in water and DMSO." Biophys J **72**(5):2032-2041.

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

van Vlijmen, H. W. and M. Karplus (1997). "PDB-based protein loop prediction:parameters for selection and methods for optimization." J Mol Biol **267**(4):975-1001.

Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." EMBO J **1**(8):945-951.

Wang, J., R. Chen and D. A. Julin (2000). "A single nuclease active site of the Escherichia coli RecBCD enzyme catalyzes single-stranded DNA degradation in both directions." J Biol Chem **275**(1):507-513.

Wang, J., R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case (2004). "Development and testing of a general amber force field." J Comput Chem **25**(9):1157-1174.

Weiserova, M., C. F. Dutta and K. Firman (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**(3):301-310.

Weiserova, M. and K. Firman (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.

Willcock, D. F., D. T. Dryden and N. E. Murray (1994). "A mutational analysis of the two motifs common to adenine methyltransferases." EMBO J **13**(16):3902-3908.

Wilson, G. G. and N. E. Murray (1991). "Restriction and modification systems." Annu Rev Genet **25**:585-627.

Yu, M., J. Souaya and D. A. Julin (1998). "Identification of the nuclease active site in the multifunctional RecBCD enzyme by creation of a chimeric enzyme." J Mol Biol **283**(4):797-808.

Zhang, X. P., R. Janke, J. Kingsley, J. Luo, C. Fasching, K. T. Ehmsen and W. D. Heyer (2013). "A conserved sequence extending motif III of the motor domain in the Snf2-family DNA translocase Rad54 is critical for ATPase activity." PLoS One **8**(12):e82184.

Zhao, R., J. Shen, M. R. Green, M. MacMorris and T. Blumenthal (2004). "Crystal structure of UAP56, a DExD/H-box protein involved in pre-mRNA splicing and mRNA export." Structure **12**(8): 1373-1381.

sinha@nh.cas.cz

Inter and intra domain interactions in the motor subunit of EcoR124I: A computational study
Ph.D. Thesis Series, 2016, No. 15.

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