

**Structural analysis of Fe(II) and 2-oxoglutarate-dependent
Dioxygenases: Escherichia coli AlkB and
Saccharomyces cerevisiae Tpa1**

Monisha Mohan

A Dissertation Submitted to
Indian Institute of Technology Hyderabad
In Partial Fulfillment of the Requirements for
The Degree of Master of Technology/ Doctor of Philosophy



भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Department of Biotechnology

June, 2015

Declaration

I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

M. Monisha.

(Signature)

M. MONISHA

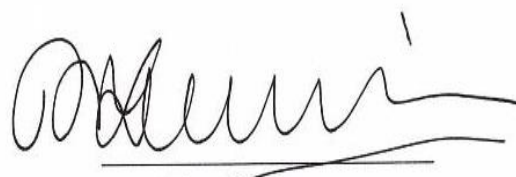
(- Student Name -)

B013M1006

(Roll No)

Approval Sheet

This thesis entitled Structural analysis of Fe(II) and 2-oxoglutarate-dependent Dioxygenases: Escherichia coli AlkB and Saccharomyces cerevisiae Tpa1 by Monisha Mohan is approved for the degree of Master of Technology/ Doctor of Philosophy from IIT Hyderabad.



Dr. Thenmalarchelvi Rathinavelan

Assistant Professor

Department of Biotechnology, IIT Hyderabad

Internal Examiner

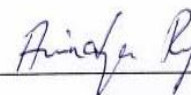


Dr. Subha Narayan Rath

Assistant Professor

Department of Biomedical Engineering, IIT Hyderabad

External Examiner



Dr. Anindya Roy

Assistant Professor

Department of Biotechnology, IIT Hyderabad

Adviser

Acknowledgements

With deep regards and profound respect, I avail the opportunity to express my deep sense of gratitude to my supervisor Dr. Anindya Roy for his inspiring guidance, constant encouragement and support throughout the period of my M.Tech thesis work.

I thank my thesis committee members Dr. Thenmalarchelvi Rathinavelan, Dr. Basant Kumar Patel and Dr. Subha Narayan Rath for their insightful comments, challenging questions and valuable suggestions.

Furthermore, I extend my gratitude to my senior lab members Naveena Kondipelli and Gururaj Shivange for their kind help and support for completion of this work.

Last but not the least, I express my deep sense of gratitude to my loving Parents, for selflessly extending their ceaseless help and moral support at all times.

Abstract

Alkylating agents induce cytotoxic DNA base adducts. Alkylation damage are usually repaired by AlkB family of enzymes, which is conserved from bacteria to human. It was first discovered in our lab that Tpa1 is involved in direct alkylation repair. Tpa1 belongs to AlkB family of Fe(II) and 2-oxoglutarate (2OG)-dependent DNA demethylase. Amino acid residues involved in 2OG and Fe (II) coordination in Tpa1 were identified using the state of the art quantum polarized ligand docking (QPLD) and molecular mechanics-generalized Born surface area (MM-GBSA) approaches. In order to design a functionally inactive tpa1 mutant without affecting the protein structure, all possible amino acid mutation using PROVEAN web server were evaluated. Result from this analysis suggested His 159 to Cys, Asp 161 to Asn, His 227 to Cys, His 237 to Cys and Arg 238 to Ala mutation. Based on this prediction, mutant tpa1 clones were generated using PCR mediated site-directed mutagenesis. As expected, Fe-binding analysis with mutant tpa1 protein did not show any characteristic absorption peak at 530nm indicating that the active site coordination of Fe and 2-OG is lost. Also, *in vivo* studies shows that mutant tpa1 along with mag1Δ is highly sensitive to MMS.

Although the structure of E.coli AlkB was solved about a decade ago, the mechanism of how AlkB is recruited to its substrate is still unknown. *In vivo* ssDNA is always coated with two major proteins namely SSB and RecA. In order to study the mechanism of loading of AlkB to ssDNA *in silico* docking programs was employed. Therefore, the molecular interaction of AlkB with ssDNA coated with SSB/RecA was studied using docking programs namely ZDOCK, PatchDock and ClusPro. Correspondingly, the interactions of AlkB with SSB/RecA was proven experimentally using *in vitro* pull down assays.

Nomenclature

1-meA – 1-methyl Adenine

3-meC – 3-methyl Cytosine

7-meG – 7-methyl Guanine

2-OG – 2-Oxo Glutarate

ACE – Atomic Contact Energy

ADT – AutoDock Tools

AlkB – Alkylation repair Protein B

BER- Base Excision Repair

CTD – C-Terminal Domain

DSBH – Double Stranded Beta Helix

FFT – Fast Fourier Transform

LigPrep – Ligand Preparation

NTD – N-Terminal Domain

MM-GBSA – Molecular Mechanics-Generalized Born Surface Area

MMS – Methyl Methane Sulfonate

MGMT – Methyl Guanine Methyl Transferase

MPE - Methyl Phosphoesters

OPLS - Optimized Potentials for Liquid Simulation

PROVEAN - PROtein Variation Analyzer

PCR – Polymerase Chain Reaction

RMSD – Root Mean Square Deviation

SAM – S-Adenosyl Methionine

SDS PAGE – Sodium Dodecyl Sulfate Poly Acrylamide Gel Electrophoresis

SSB - Single-stranded DNA-binding proteins

TLS – Trans-Lesion synthesis

Tpa – Termination and polyadenylation associated

Contents

Declaration.....	Error! Bookmark not defined.
Approval Sheet	Error! Bookmark not defined.
Acknowledgements.....	iv
Abstract.....	v
Nomenclature	vi
1 Ligand Docking and Binding Site Mutation Prediction of AlkB Homolog of Saccharomyces cerevisiae	1
1.1 Introduction.....	1
1.1.1 Alkylation damage to DNA.....	1
1.1.2 DNA is exposed to the endogenous alkylation agents.....	2
1.1.3 Experimentally used DNA damaging alkylating agents.....	2
1.1.4 Cellular defense against alkyl-base lesions	3
1.1.5 Structure of E.coli AlkB reveals molecular mechanism of DNA repair	3
1.1.6 Background of DNA repair research in the lab	4
1.1.7 Structure of Tpa1 reveals Fe (II)/2OG-dependent dioxygenase domain.....	5
1.2 Objectives	7
1.3 Materials and methods	8
1.3.1 Agarose gel electrophoresis.....	8
1.3.2 Preparation of competent E.Coli cells	9
1.3.3 Transformation of competent E. coli cells.....	9
1.3.4 Plasmid DNA Isolation from mini bacterial culture.....	10
1.3.5 SDS-PAGE.....	11
1.3.6 Other solutions	12
1.4 Design of Tpa1 mutations by molecular docking	13
1.4.1 Protein Structure Preparation	13
1.4.2 Ligand preparation.....	13
1.4.3 Autodock 4.2 Methodology	13
1.4.4 GLIDE QPLD (QM-polarized ligand docking):	13
1.4.5 Analysis and visualization of docking simulation results.....	13
1.4.6 Validation of the docked poses/ molecular docking.....	14
1.4.7 <i>In silico</i> Site Directed mutagenesis of Tpa1	14

1.5	Site-directed mutagenesis of Tpa1D	14
1.5.1	Construction of mutant Tpa1	14
1.5.2	Tpa1 Expression Analysis of the recombinant mutant proteins	16
1.5.3	Purification of GST-tag mutant Tpa1 Proteins.....	Error! Bookmark not defined. 16
1.6	Results and Discussions.....	18
1.6.1	Tpa1 is structurally similar to AlkB	18
1.6.2	First attempt for making mutant Tpa1	19
1.6.3	Molecular Docking WT Tpa1 with Autodock 4.2.2.....	25
1.6.4	Molecular Docking of WT Tpa1 with GLIDE QPLD.....	27
1.6.5	Prediction of amino acid substitution by PROVEAN algorithm	28
1.6.6	<i>In silico</i> mutagenesis with PyMol:.....	29
1.6.7	Molecular Docking of mutant Tpa1 with AutoDock	29
1.6.8	Molecular Docking of mutant Tpa1 with GLIDE QPLD	34
2	Protein Docking and Prediction of Molecular Interactions of E. coli AlkB	37
2.1	Introduction.....	37
2.1.1	Role of AlkB in the repair of damaged ssDNA	38
2.1.2	Mechanism of repair of damaged base by AlkB protein	38
2.1.3	Structure of AlkB protein	40
2.1.4	Role of Single-stranded DNA-binding protein.....	42
2.1.5	Structure of Single Strand DNA Binding (SSB) protein	42
2.1.6	Structure of RecA protein	45
2.2	Materials and methods	47
2.2.1	Protein Structure Preparation	47
2.2.2	Docking analysis using PATCHDOCK	47
2.2.3	Docking analysis using ZDOCK.....	48
2.2.4	Docking analysis using CLUSPRO.....	48
2.2.5	Analysis and visualization of docking simulation results.....	48
2.3	Objective.....	49
2.4	Results and Discussion	50
2.4.1	Docking of AlkB with ssDNA wrapped SSB tetramer	50
2.4.2	Docking of AlkB with ssDNA bound RecA filament	58
	References.....	65

PART-1

1. Ligand Docking and Binding Site Mutation Prediction of AlkB Homolog of *Saccharomyces cerevisiae*

1.1 Introduction

1.1.1 Alkylation damage to DNA: Alkylating agents can introduce methyl or ethyl groups at all of the available nitrogen and oxygen atoms in DNA bases, producing eleven different types of base modifications. Among these, 7-methylguanine, 3-methyladenine (3meA) and O6-methylguanine (O6meG), are specific to the double stranded DNA. Whilst 7-methylguanine is relatively harmless, 3meA has a strong toxic effect. The methyl group of 3meA protrudes into the minor groove of the double helix which is normally free of methyl groups (the methyl groups of thymine and 5 methyl cytosine are in the major groove). Here, 3meA efficiently blocks RNA- and most DNA polymerases resulting in a strong cytotoxic but low mutagenic effect. The low mutagenicity observed may be due to translesion synthesis by nonreplicative DNA polymerases. O6meG is highly mutagenic and directly demethylated by O6-methylguanine-DNA methyltransferase (MGMT). This protein transfers the methyl group to a Cys residue within the active site resulting in self-inactivation.

In single stranded DNA 1-methyladenine (1-meA) and 3- methylcytosine (3-meC) lesions are mostly observed as all the other nitrogen lone pairs required for nucleophilic attack are occupied with hydrogen bonding in double stranded DNA (1). These lesions are strong toxic but weak mutagenic as they are unable to base pair but able to block DNA replication (1,2). The low mutagenicity observed may be due to translesion synthesis by nonreplicative DNA polymerases.

1.1.2 DNA is exposed to the endogenous alkylation agents: Endogenous sources of alkylation damage come from many types of reactive molecules inside cells. The enzyme cofactor S-adenosylmethionine is the methyl donor in many biomethylation pathways and is present in cells in a concentrations ranging from 20 μM to 50 μM . This molecule can methylate DNA at a rate characteristic of weak alkylating agents (3). The peroxidation of lipids in the cell membranes by free radical byproducts of cellular metabolism can create alkylating agents that damage DNA. The main lesions produced by this mechanism are ethenoadenine (ϵA) and ethenocytosine (ϵC).

1.1.3 Experimentally used DNA damaging alkylating agents: Cells are always exposed to DNA damaging alkylating agents that occur in the environment and endogenously generated byproducts of cellular oxidative metabolism which leads to the damage in genome. Some alkylating agents such as naphthyridinomycin and cyanocycline A are also used in cancer chemotherapy (4). Simple methylating agents, such MMS, methylates double-stranded DNA and generate 7-methylguanine (7meG) and 3-methyladenine (3meA) (5).

1.1.4 Cellular defense against alkyl-base lesions: Specialized damage-tolerant DNA polymerases, executes trans-lesion synthesis (TLS) when replicative polymerases are stalled at these lesions (6). In *S. cerevisiae*, TLS is carried out by the Pol η (Rad30), Pol ζ (Rev3, Rev7), and Rev1 polymerases, all of which have human homologs (7). Yeast Pol η is dedicated to the repair of UV-induced cyclobutane-pyrimidine dimers in an error-free manner (8). In contrast, Pol ζ , in cooperation with Rev1, participates in error-prone TLS across lesions produced by variety of DNA-damaging agents, including MMS (9). In yeast, DNA glycosylase Mag1 specifically removes 3meA. Following Mag1-mediated removal of damaged base, Apn1 apurinic/apyrimidinic endonuclease

cleaves the DNA strand at the abasic site for subsequent repair of the single strand break by base excision repair (BER) pathway. The enzyme that directly repairs 1-meA and 3-meC is known as alkylation repair protein-B (AlkB) in *E. coli* (10,11). AlkB deficient *E. coli* cells accumulate alkylated lesions and are hypersensitive to alkylating agents (12). AlkB catalyzed demethylation reaction is coupled to the oxidative decarboxylation of 2OG to succinate and CO₂ resulting removal of methyl group from 1-meA and 3-meC. The methyl group is hydroxylated and spontaneously released as formaldehyde (13, 14).

1.1.5 Structure of E.coli AlkB reveals molecular mechanism of DNA repair: The *E. coli* AlkB protein utilizes an oxidative dealkylation mechanism of removal of DNA damage. AlkB is a member of the large Fe(II) and 2OG-dependent dioxygenases family and show similar conserved features like conserved HxD_nH that coordinate the 2OG and iron (Fe(II)) and catalytic core consisting of double strand β-helix (DSBH) fold (15, 16), also known as a jelly-roll fold or double Greek key motif. The Fe(II) binding site of 2OG oxygenases is located at the more open end of the DSBH core. Three residues forming a conserved H_xD/E. . . H motif, and coordinate the Fe(II) . The H_xD/E of the motif is located at the end of the second strand of the DSBH and the distal histidine of the motif on the seventh strand of the DSBH. The same metal coordinating triad is conserved in all characterized Fe(II)/2OG-dependent dioxygenases.

Recently available structures of the un-liganded AlkB (pdbid: 3KHB) and the AlkB-ssDNA complex containing 1-meG (pdbid: 3KHC) reveals several interesting mechanistic details. The protein utilizes both hydrophobic interactions and hydrogen bonding that stretch across five nucleotides to maintain the protein-DNA complex. The phosphodiester backbone of the ssDNA binds in an electropositive DNA binding groove

that is created by a T51-Y55 DNA binding loop, S129, and K127. Alkyl-lesions are flipped into the substrate binding pocket where the metal binding residues (H131, D133, and H187) and substrate binding residues come close to each other. The substrate base is bound in the active site pocket formed by hydrophobic stacking interaction of W69 and H131. These interactions also provide the stability to maintain the flipped conformation of 1-meG in the substrate binding pocket. The substrate base is bound in the active site pocket formed by hydrophobic stacking interaction of W69 and H131. The methyl group of the 1-meG is adjacent to the bound metal ion and properly positioned for oxidative repair. The hydroxyl group of T51 and backbone amide of G53 form direct hydrogen bonding interactions to the phosphate of the nucleotide 5'-adjacent the 1-meG base. Amino acids Y76, K127 and S129 provide additional contacts with the phosphodiester backbone, while the side chain of Y55 forms a hydrophobic packing interaction with the ribose sugar. The general interactions between the protein and phosphodiester backbone or methylated base of the nucleic acid provide few - opportunities for distinguishing between ribonucleotides and deoxyribonucleotide substrates, which is consistent with the ability of AlkB to function on both DNA and RNA.

1.1.6 Background of DNA repair research in the lab: Homologs of AlkB were identified across species ranging from bacteria to human (17, 18) except *Saccharomyces cerevisiae* (19). It was reported earlier that two genes of *S. cerevisiae* namely *YFW1* and *YFW12* could complement deficiency of AlkB in *E.coli* (20). However, *YFW1* is an endoplasmic reticulum membrane protein and *YFW12* is secreted-sterol binding protein, and they share no sequence homology with AlkB or any other Fe(II) and 2OG-dependent dioxygenases (21, 22) and, therefore, could not be considered AlkB

homolog. No genetic interactions were reported. Although functional homolog of AlkB remained unknown in *S. cerevisiae*, search for the dioxygenase domain containing proteins in budding yeast revealed that an uncharacterized ORF named *YER049W* had the characteristic dioxygenase domain (23). Later the gene product of *YER049W* was renamed as ‘termination and polyadenylation protein’ (Tpa1) as it was found to be associated with eRF1, eRF3 as well as polyA binding protein (PABP) within the mRNA ribonucleoprotein complex (24). TPA1 deletion in yeast resulted decrease of translation termination efficacy and an increase in mRNAs stability (25). A recent study demonstrated that Tpa1 probably function as prolylhydroxylase responsible for hydroxylation 40S ribosomal subunit protein (27). However, none of these studies provided any direct evidence for prolylhydroxylase enzymatic activity using purified Tpa1 (25–27). Our lab initiated study on Tpa1 in response to the findings that Tpa1 is the only *S. cerevisiae* protein that belongs to Fe(II) and 2OG-dependent dioxygenase superfamily of proteins which also includes AlkB (23). Interestingly, genetic screen in yeast deletion mutants revealed that TPA1 deletion caused mild MMS sensitivity (28), suggesting that this protein might be involved in repair of DNA alkylation damage. Our lab established that purified recombinant Tpa1 catalyzes the oxidative demethylation of methylated DNA and promote survival of MMS sensitive *E. coli alkB* mutant cells. Furthermore, we demonstrated genetic interaction between Tpa1, DNA glycosylase Mag1 and TLS polymerases Polζ (Rev3) in *S. cerevisiae*. Our lab also showed that Mag1 appears to have a synergistic relationship with Tpa1, because *tpa1Δmag1Δ* double mutant showed exacerbated phenotype. Most notably, we have uncovered a remarkable synergism among Mag1, Tpa1, and TLS polymerases Polζ (Rev3) in protecting against methylation damage, as indicated by the inability of the

tpa1Δmag1Δrev3Δ triple mutant cells to recover from extremely low level of MMS-induced methylation damage.

1.1.7 Structure of Tpa1 reveals Fe (II)/2OG-dependent dioxygenase domain: Tpa1 is composed of two stacked domains forming a cylinder of 90 Å length and 45 Å diameter. The N-terminal domain encompasses amino acids 26–260 (residues 1–25 are not visible in the electron density maps), whereas the C-terminal domain encompasses amino acid 293–635 (residues 636–644 are not visible in the electron density maps). Both domains are packed together and connected by a linker (amino acids 261–292). The DSBH-fold is shared by a wide variety of Fe(II)/2OG-dependent dioxygenases. Structural alignments with Tpa1 NTD reveal similarity with several Fe(II)/2OG-dependent dioxygenases. Analysis of the amino acid side chains strictly conserved between Tpa1-NTD and other Fe(II)/2OG-dependent dioxygenases reveals that the H_xD ...H triad (positions 159, 161, and 227 in Tpa1, respectively) superposes perfectly with the characteristic H_xD ...H triad of Fe(II) binding motif. Just like other typical DSBH enzymes, these three residues form a triad, leaving three coordination sites on the octahedral Fe (II) center for the binding of 2OG and dioxygen.

1.2 Objectives:

- 1. Designing active site mutation of Tpa1 by molecular docking**
- 2. Site-directed mutagenesis of Tpa1**
- 3. Phenotypic analysis of Tpa1 mutation**

1.3 Materials and methods

General Methods

1.3.1 Agarose gel electrophoresis: Horizontal agarose gel (0.8-1.0%) electrophoresis were routinely performed to separate DNA fragments of various lengths (ranging from 100bp-10kb). Appropriate amount of agarose (SeaKem®LE, LONZA, Cat. no. 50005) was dissolved in 1X TAE buffer (40mM Tris-acetate; 1mM EDTA pH 8.0) by heating in microwave oven. After cooling, Ethidium bromide was added of concentration 1µg/ml into the gel solution and then poured into gel a mould and a comb was inserted to generate wells. After 30-45 minutes, when the gel got solidified completely, comb was removed gently and gel mounted into electrophoresis chamber containing 1X TAE (40mM Tris-acetate; 1mM EDTA pH 8.0). DNA samples and size marker were mixed with appropriate volume of 1X DNA loading dye (6x stock: 0.25% bromophenol blue; 0.25 xylene cyanol FF; 30% v/v glycerol, (Cat.no:R0611, Thermo scientific). Electrophoresis were generally carried out at 80 volt till the xylene cynol dye migrated to distance of half of the gel. DNA bands were visualized in gene documentation system from SynGene (Model no: Chemi XR5, S.No DR4V2/2355). 1kb ladder (Cat no: SM0311 Thermo scientific) for Tpa1, linearized plasmid DNA were used for fragment size determination.

1.3.2 Preparation of competent *E.Coli* cells: Competent *E. coli* cells of DH5α and BL21 (DE3) pLysS, were prepared described by Alexander by using MnCl₂ and CaCl₂. A sterilized inoculating loop was used to streak *E.coli* strains (DH5α and BL21 (DE3) pLysS) directly from frozen glycerol stock onto an LB agar plate containing no antibiotic. Plate was incubated for 16 hours at 37°C. A single colony was picked and

inoculated in 5 ml of LB medium for overnight in a shaking incubator at 200 rpm. 1ml of this overnight grown culture was inoculated in a pre-warmed 100 ml of LB medium (1% tryptone, 0.5% yeast extract) prepared in 250ml Erlenmeyer flask. Inoculated culture was grown for 4hrs at 30°C, 100 rpm in shaking incubator under monitoring culture growth by measuring OD₆₀₀ spectrophotometer every 20 minutes. When culture had reached OD₆₀₀ of 0.35 it was taken out and incubated on ice for 1hr. Cells were harvested by centrifugation at 4000 rpm for 15 min at 4°C. The supernatant was discarded and the cell pellet was resuspended in 12 ml of acid salt buffer-A (ASB-A, sodium acetate-40 mM, CaCl₂-100 mM, MnCl₂-70 mM, and pH 5.5) and incubated on ice for 1 hour. The ASB treated cells were then pelleted by centrifugation carried out at 3500 rpm for 15 min at 4°C and resuspended in 4ml of ASB-B (Sodium acetate-40 mM, CaCl₂-100 mM, MnCl₂-70 mM, 15% glycerol, pH 5.5) and were stored in Eppendorf tubes (aliquots of 50µl) at -86 °C for future use.

1.3.3 Transformation of competent E. coli cells:

(I) *Transformation of competent DH5α cells:* DH5α competent E.coli cells were prepared as described earlier. 10µl of a ligation mixture were added to 50µl aliquots of competent DH5 α cells and incubated on ice for 30 minutes. After that heat shock was given at 42°C for 30 seconds, cells were again briefly incubated on ice for 5 minutes and transformed competent cells were plated on LB- agar plates containing 100µg/ml ampicillin. Plates were incubated at 37°C overnight.

(II) *Transformation of BL21 (DE3) pLysS competent cells:* BL21 (DE3) pLysS competent E.coli cells were prepared as described earlier. 4µl of a plasmid DNA was added to 50 µl aliquots of cells and the mixture was incubated on ice for 10 minutes. After a 42°C heat shock for 30 seconds, the cells were briefly incubated on ice for 5

minutes. After this incubation, entire transformation mixture was plated on LB-agar plates containing 100µg/ml ampicillin. Plates were incubated at 37°C overnight for growth of cells.

1.3.4 Plasmid DNA Isolation from mini bacterial culture: A single colony was picked out from the plate containing colonies of transformed cells and inoculated in 5ml of LB medium containing ampicillin. The culture was incubated in a shaking incubator at 37°C, 200 rpm for overnight. 2ml of this culture was used for isolation of plasmid DNA using Plasmid Mini-Prep kits (Fermentas Life Sciences, cat. no K0502) following the instructions recommended by the manufacturer.

1.3.5 SDS-PAGE: Recombinant proteins were analyzed through SDS-PAGE. This was carried using discontinuous buffer system as described by Laemmli using Biorad gel electrophoresis apparatus (Cat. no: 165-8001). Throughout this study 1.5 mm thick 12% polyacrylamide (30:0.8, acrylamide to bisacrylamide ratio) gel containing 0.1% SDS was used for the separation of proteins. For some application Biorad precast gel were also used (Cat. no: 456-1086). The protein samples were mixed with an equal volume of 3X sample buffer (2.4 ml 1M Tris-HCl pH 6.5, 3 ml 20% SDS, 3 ml 100% Glycerol, 1.6 ml β-mercaptoethanol, 0.006 g Bromophenol blue) and heated at 100 °C for 10 min and were loaded into the wells of precast polyacrylamide gel. Electrophoresis was performed at a constant voltage of 100 or 120V by diluting 10x buffer (Tris base 30.3g, Glycine 144g, SDS 10g in 1000ml). After the electrophoresis, the gel was stained with coomassie brilliant blue (0.4% w/v Coomassie blue R250, 30% v/v Methanol, 10 % Acetic acid). Spectra multicolor broad range protein ladder was used to confirm the size of protein. Composition of 12.5% SDS-PAGE gel is given below:

Table7. SDS-PAGE gel composition

Resolving Gel (12.5 %)		Stacking Gel (4 %)	
30% Acrylamide	2.00 ml	30% Acrylamide	0.65 ml
Tris-HCl, pH 8.8, 1.5M	1.25 ml	Tris-HCl, pH 6.8, 0.5M	1.25 ml
H ₂ O	1.70 ml	H ₂ O	3.05 ml
SDS (10%)	50µl	SDS (10%)	50µl
TEMED	20µl	TEMED	10µl
APS	45µl	APS	25 µl

Media and Solutions: De-ionized water was used for all buffers solutions and media.

Media	Composition
LB media	1% Tryptone w/v, 0.5% w/v Yeast Extract
TB media	12g Tryptone , 24 g Yeast Extract 4ml glycerol in 1L
SOC media	2.0% Tryptone w/v, 0.5 % w/v Yeast extract.

1.3.6 Other solutions:

Isopropyl β-D-1-thiogalactopyranoside (1M): IPTG 4.7g, H₂O 20ml.

20% SDS: SDS (20g), H₂O (80ml). Make final volume 100 ml with stirring.

20% Ammonium per Sulphate: APS (0.2g), H₂O (0.8 ml).

General buffers used in the study:

Buffers	Composition
TE (Tris-EDTA) 50X	1mM Tris-HCl pH 8.0, 0.1 mM EDTA pH 8.0. A 10X stock solution was routinely used to prepare 1X TE.
SDS-PAGE Resolving gel buffer	1.5 M Tris-HCl, (for Resolving Gel) ,pH 8.8
SDS -PAGE Stacking gel buffer	0.5 M Tris-HCl (for stacking gel), pH 6.5
10X SDS-PAGE gel running buffer	Tris base 30.3g, Glycine 144g, SDS 10g (add last) , makefinal volume to 1000ml.
1X Gradient gel running buffer	25mM Tris Base, 192mM Glycine, 0.1 % (w/v) SDS, methanol 20%.
3X SDS-PAGE loading dye (10ml)	1M Tris-Cl, pH 6.5 (2.4 ml), 20% SDS (3 ml), Glycerol (100%) (3 ml), β-mercaptoethanol (1.6 ml), Bromophenol blue (0.006g).
SDS-PAGE staining solution.	Coomassie blue R250- 0.4 % w/v, Methanol- 30 % v/v, Acetic acid 10 %v/v
SDS-PAGE destaining solution	Methanol- 30 % v/v, Acetic acid 10 % v/v.

1.4 Design of Tpa1 mutations by molecular docking

The atomic coordinates of the Tpa1 protein (PDB ID: 3KT7) was obtained from Protein Data Bank. The interaction of ligand with native protein was monitored using Discovery Studio Visualizer 2.5 (28).

1.4.1 Protein Structure Preparation

The Protein Preparation Wizard (PrepWizard) of the Schrodinger suite (Schrödinger, LLC, New York, NY, 2014) was implemented to preprocess, which assigns the bond orders and adds hydrogen to the protein structure. The protein was optimized using PROPKA, which predicts the pKa of the ligand for accurate molecular docking (28). The optimized protein was then energy minimized using Optimized Potentials for Liquid Simulation (OPLS_2005, Schrödinger, New York, NY, and USA) force field.

1.4.2 Ligand preparation

The ligand molecule was prepared using LigPrep wizard of the Schrodinger suite. The ligand was ionized using EPIK at biologically pH of 7 +/- 2 and its structure was minimized using OPLS_2005 force field (29).

1.4.3 Autodock 4.2 Methodology

The prepared ligand was docked to the protein following the standard protocol involving flexible ligand and rigid protein. The Kollman charges were added to the prepared Tpa1 protein using AutoDock tools (ADT). The grid maps were generated using Autogrid. Lamarckian genetic algorithm methodology was employed for docking simulations in AutoDock 4.2 (30). The ligand docking involved 10 independent runs

per genetic algorithm. The population size of 150 with maximum number of evaluation of 250000 and maximum number of generation of 27000.

1.4.4 GLIDE

The docking accuracy can be achieved through GLIDE QPLD which considers the polarization of ligands for docking with the protein (31). The binding box with size 10ÅX10ÅX10Å was centered on the ligand bound to the protein. The default setting was used for the generation of the grid. The protocol involves docking of the ligand to the rigid protein file, which generates top five poses. The ligands with QM/MM modified charges were relocked and then Prime MM-GBSA method (Molecular Mechanics-Generalized Born Surface Area) was used to calculate the binding free energy of the ligand (31).

1.4.5 Analysis and visualization of docking simulation results

Docking analysis was performed to identify the possible conformations and orientations of the ligand at the binding site. The best conformation of the ligand was selected based on two criteria: the interactions with the active site residues and its thermodynamic stability. The best poses were analyzed for non-covalent interactions using Discovery Studio Visualizer 2.5 and PyMol.

1.4.6 Validation of the docked poses/ molecular docking

To know about the reliability of the molecular docking, the co-crystallized AKG was removed and again re-docked on to the Tpa1 NTD binding site. The RMSD was calculated using superimposition option in the Schrodinger suite.

1.4.7 *In silico* Site Directed mutagenesis of Tpa1

In order to generate the mutant tpa1, substitution mutation was introduced in *silico* using site directed mutagenesis wizard in PyMol. To identify the amino acid substituent that would affect the function of the protein drastically PROVEAN (PROtein Variation ANalyzer) algorithm was employed (32). H159C, D161N, H227C, H237C and H238C was introduced in mutant Tpa1 protein. FoldX algorithm was used to make sure that the mutations did not affect the overall stability of protein (33).

1.5 Site-directed mutagenesis of Tpa1

1.5.1 Construction of mutant Tpa1:

The construct pGEX-6P-1 was used for the bacterial expression of N-terminal GST tagged mutant Tpa1. Two rounds of PCR protocol was followed for generating mutant Tpa1 using the two different primer pairs. In the first round two mutations i.e. H159C, D161N megaoligo was generated using pGEX-TPA1 as template. The primer pairs used are Primer1 (Mut-F-H159C, D161N) GGG ATG TCA CTT GTT GAC TTG TGA TAA CGT TAT TGG CTC TAG AAG GAT TAG and Primer2 (GST-TPA1-Anti) AAA AGT CGA CAA GCT TTT ACG CTT CAT CTT CCT GAC CAT CTTC. The first round PCR product is digested with HindIII, gel eluted and then used as a megaoligo for the second round mutagenesis PCR. After PCR, the PCR product was digested with

DpnI and then transformed into DH5 α . Plasmid thus obtained was used as template for generation of other 3 mutations in the second round of PCR. The other three mutations which include H227C, H237C, R238A, were generated as megaoligo. The primer pairs used are: Primer3 (Mut-F-H159C,D161N) GGG ATG TCA CTT GTT GAC **TTG TGA TAA CGT TAT TGG CTC TAG AAG GATTA** and Primer4 (Mut-R-H227C,H237C,R238A) GGT ACC ATC CTT GAA TAG ATA **ATG CGC ACT TAT CAA CTT TGA CTT CGT CGA CAT CGC AGA AAG AGA AAC CCGG**. Sequences in bold represents the mutation site for Histidine159, Aspartate161, Histidine227, Histidine237, Arginine238 of Tpa1 to Cysteine, Asparagine, Cysteine, Cysteine and Alanine respectively, along with generation of SalI restriction site(underlined) for screening the mutant. All PCR amplifications were performed using the following conditions with Phusion Hifidelity DNA polymerase; initial denaturation at 98°C for 30 sec (1 cycle); denaturation at 98°C for 10 sec; annealing at 57°C for 30 sec and extension at 72°C for 60 sec for 32 cycles and final extension at 72°C for 10 mins. After the PCR, the product is digested with DpnI (New England Biolabs) and then used for transformation. Obtained colonies were screened for mutant clones by the release of insert using confirmation of the mutant clone. Plasmid DNA isolation was carried out as described before.

Table 4. Primers used for site directed mutagenesis PCR.

Target Residues	Name of the Primer	Sequence (5'-3')
Tpa1 Mutagenesis	Mut-F-H159C,D161N	GGGATGTCACCTTGTTGACTTGTGATAACGTTATTGCCTCTAGAAGGATTAG
	Mut-F-I171A	GGCTCTAAAGGATTAGTTTCGCTTTAGCCCTACCAATCCCCGACAGAAAA TGGAAATCGC
	Mut-R-H227C	GGTACCATCCTTGAATAGATAATGCGCACTTATCAACTTTGACTTCGTCGA CATCGCAGAAAGAGAAACCCGG

1.5.2 Expression Analysis of the recombinant mutant proteins: *E.coli* BL21 (DE3) pLysS cells transformed with recombinant plasmids were grown and induced for protein expression as described by Studier *et al.* A single colony was inoculated in 6 ml of Luria-Bertani medium containing 100µg/ml ampicillin and grown for 4 hrs at 37°C in a screw-cap inoculation vial. From this culture 2ml culture was recovered as a reference for uninduced *E.coli* extract and the remaining 4ml culture was induced with 1mM (final concentration) isopropyl-β-D-thiogalactopyranoside (IPTG). Following induction the culture was further grown for 4 h at 37°C and 1.5ml culture was recovered from the induced sample as a reference for induced *E.coli* extract and 750 µl was recovered from uninduced culture. The IPTG-induced and uninduced cells were harvested and resuspended in 100µl protein extraction buffer (50 mM Tris pH 8.8, 250 mM NaCl, 0.05% Triton-X) and lysed by sonication. From the sonicated samples, 60µl were added to 30µl of 3x sample loading buffer, boiled for 10 min and were used for SDS-PAGE analysis in 12% gel.

1.5.3 Purification of GST-tag mutant Tpa1 Proteins: GST-tag Tpa1 was checked for overexpression and purified by GST-glutathione affinity purification method. Recombinant plasmid DNA showing over expression of recombinant GST fusion protein in expression analysis were transformed into BL21 (DE3) pLysS cells. 270ml

LB broth with amp100µg/ml was grown for preinoculum, and inoculated to 2.7L of TB containing amp of concentration100µg/ml. Following induction, cells were pelleted at 8000 rpm, 15min, 4°C. Pellet was stored at -86°C. Pellet was resuspended in 10ml of Extraction buffer containing the protease inhibitor tablet “complete” (Roche, REF 11836170001). Dissolved pellet was sonicated. A 20µl aliquot of cell lysate was store at -86°C Cell lysate was centrifuged at 14000 rpm, 20 min at 4°C to separate soluble and insoluble fraction. Supernatant was collected in 50ml falcon tube. A 20µl aliquot of supernatant and insoluble fraction was stored for SDS-PAGE analysis to identify the presence of fusion protein in the fractions. 800µl of glutathione sepharose 4B slurry was added to the supernatant and kept it for binding to at 4°C cold chamber on rotary shaker for 2hrs. The tube was centrifuged at 500g for 5min at 4°C to sediment the protein bound glutathione matrix. Supernatant was collected as flow through (unbound fraction) in a 50 ml falcon tube. A 20µl aliquot of flow through was also collected for SDS-PAGE analysis. Beads settled on the bottom of tube were washed with 10ml wash buffer for 10 min on rotary shaker at 4°C and then centrifuge at 500g/4 min/4°C. The unbound fraction was collected by gentle pipetting. Washing step was repeated three times. Glutathione beads bound to protein were transferred to 2ml tube. 1.0 ml of elution buffer (50mM Tris (pH 8.0), 150mM NaCl, 0.05% TritonX, 10mM reduced glutathione, 10% glycerol) was added to beads and kept on rotary shaker for 25 min. To elute the protein, beads were centrifuged at 500g/5min 4°C and supernatant was collected. This process was repeated two more times and total of 3 elutions of 1ml each were collected. Aliquot 5µl of each elution was analysed by SDS-PAGE. Eluted proteins were stored at -20°C.

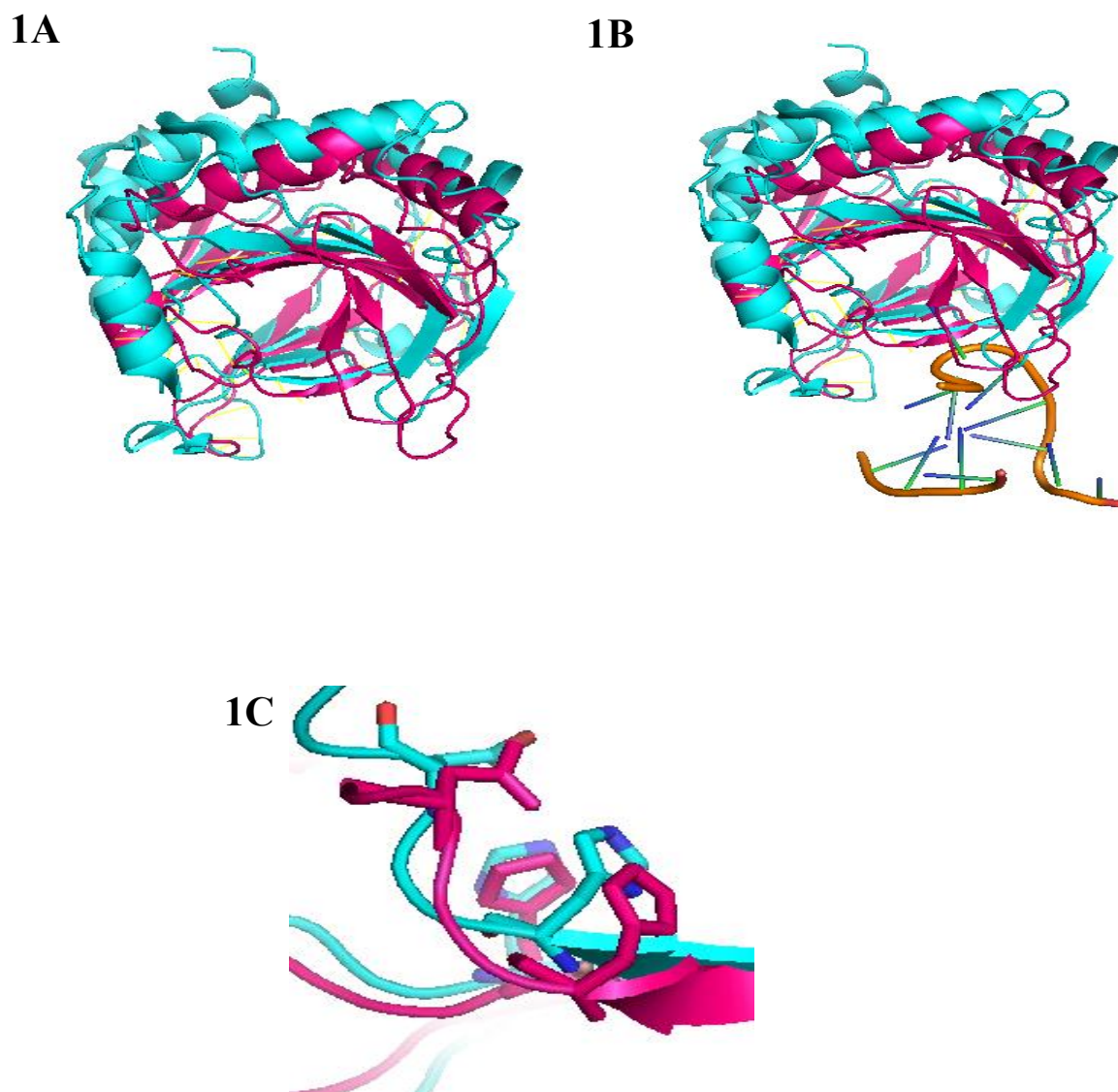
1.6 Results and Discussions

Analysis of Tpa1 structure and identification of key residues involved in cofactor binding

1.6.1 Tpa1 is structurally similar to AlkB: To examine similarity between Tpa1 and AlkB, we superimposed the *S. cerevisiae* Tpa1 enzyme structure onto the *E. coli* structure. This protein has 2 domains namely, N-Terminal Domain (NTD) and C-Terminal Domain (CTD). X-ray crystal structure reveals that both the domains have the characteristic DSBH fold and the r.m.s.d. between NTD and CTD is 2.8 Å for 180C α atom pairs. Despite this similarity Fe (III) and 2-OG was present at the active site of NTD only. Tpa1 NTD superimposes onto the structure of AlkB (PDB 3I3Q) with an r.m.s. deviation of 2.6Å over 135–140 C α atoms. The overall similarity is particularly evident in the central DSBH/jelly-roll core, where the iron-binding triad as well as 2OG adapts nearly identical geometry in the two enzymes. Structure-based sequence alignment reveals relatively low sequence conservation with 15.5% overall identity (28.2% similarity) between Tpa1 and AlkB. Despite the existing information regarding AlkB:DNA recognition and activity, how Tpa1 might recognize single-stranded DNA (ssDNA) is an open question. In order to understand the active site architecture that defines substrate specificity of Tpa1 we decided to carry out structure function analysis of Tpa1 activity.

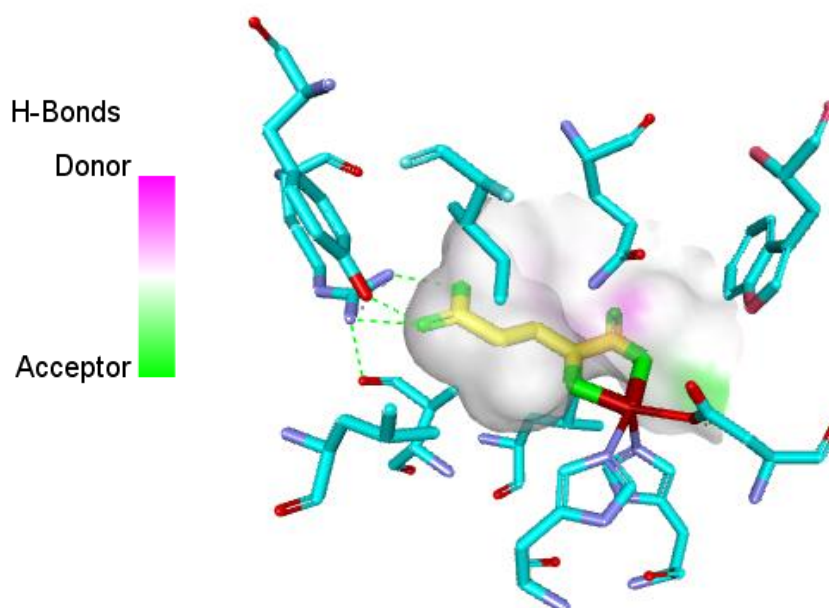
1.6.2 First attempt for making mutant Tpa1: The ternary complex of Tpa1 with both 2-OG and Fe was used to study the structure and to analyse the active site residues of Tpa1. Applying the PyMol align option we identified that the characteristic H₁₃₁D₁₃₃...H₁₈₅ triad of the AlkB superimposed well with Tpa1 H₁₃₁D₁₃₃...H₂₂₇ with r.m.s.d of 0.280Å (**Figure 1**). The Tpa1 active site was further analysed using Discovery Studio visualizer 4.0 to study the hydrogen bonding (**Figure 2**), hydrophobic (**Figure 3**) and electrostatic interactions (**Figure 4**). It was observed that the Tpa1 protein interacts with Fe through H159, D161 and H227. The Fe coordinates with 2-OG which in turn interacts with I171, Y173, V229, R238 and S240 residues. To validate this interaction a mutant Tpa1 was generated where H159, D161, H227, R238 and S240 were mutated to Alanine. However, functional analysis with this Alanine mutant revealed that there was no loss of activity (Gururaj Shivange, Navena Kodipelli, Anindya Roy, personal communication). Therefore it was very important to design functionally inactive Tpa1 mutant.

Figure1:



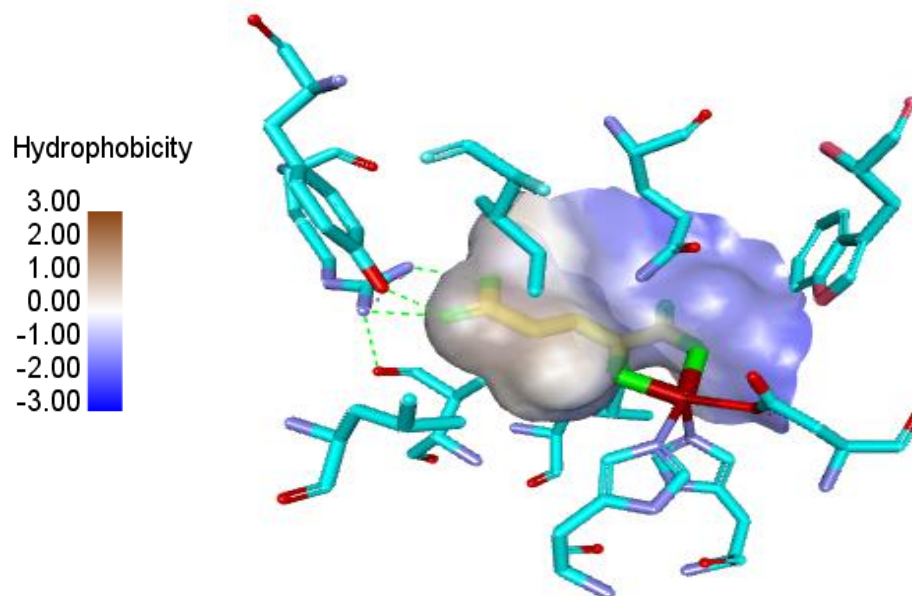
Structural comparison of Tpa1 and AlkB: Structural alignment of Tpa1 with AlkB in the absence (1A) and presence of DNA (1B). Superimposition of the HDH triad, a characteristic motif of Fe (II) - dependent Dioxygenases (1C). The active site triad superimposed with RMSD of 0.208Å. The magenta and cyan ribbons indicate the secondary structure of AlkB and Tpa1 proteins respectively. The structures were aligned using PyMol.

Figure2



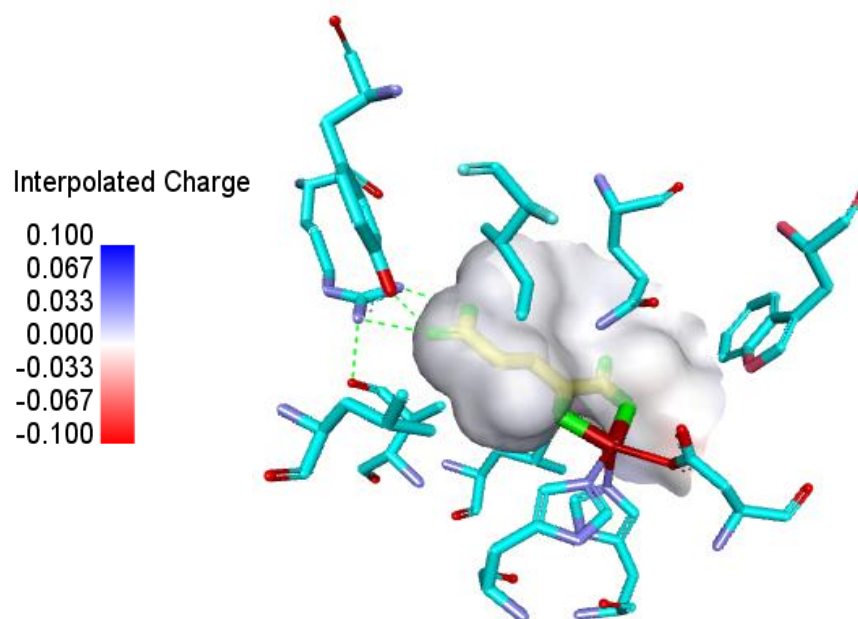
Analysis of Hydrogen bonding pattern of Tpa1 with ligand: The ternary complex of Tpa1 with cofactor Fe and co-substrate 2-OG (PDB ID: 3KT7) was used for the analysis. The residues R238 and Y173 in the active site of Tpa1 protein forms hydrogen bond with 2-OG. The Hydrogen bond donor and acceptors are shown in pink and green color surface around the ligand. The interaction of 2-OG (carbons colored in yellow) with Fe (red ball) and Tpa1 (amino acid residues in cyan sticks) was studied using Discovery studio visualizer. The hydrogen bond distance parameters was set as $D - A < 3.4\text{\AA}$ and the bond angle criteria as $90 < D - H \dots A < 180$.

Figure 3



Analysis of Hydrophobic interactions of Tpa1 with ligand: The Tpa1 structure in complex with cofactor Fe and co-substrate 2-OG (PDB ID: 3KT7) was used for the identification of the hydrophobic residues in the active site. The ligand 2-OG is coordinated with Fe atom (shown as maroon sphere). The residues Y150, Y173, I171, L156, L189 and W244 shows hydrophobic interactions with 2-OG. The surface map around the ligand indicates the level of hydrophobicity. The brown regions is the hydrophobic patch and the blue region is hydrophilic which interacts with the polar amino acids such as D161 and Q242 of the Tpa1. The interaction of 2-OG (carbons colored in yellow) with Fe and Tpa1 (amino acid residues in cyan sticks) was monitored using Discovery studio visualizer.

Figure 4



Analysis of Electrostatic interactions of Tpa1 with ligand: The ternary complex of Tpa1 with cofactor Fe and co-substrate 2-OG (PDB ID: 3KT7) was used for the analysis. The interaction of 2-OG (carbons colored in yellow) with Fe (red ball) with Tpa1 (amino acid residues in cyan sticks) was studied using Discovery studio visualizer. The ligand 2-OG is coordinated with Fe atom shown in maroon sphere. The surface map around the ligand is shown in white indicating that the electrostatic interaction of the ligand with the protein is almost neutral. The bar on the left indicates the interpolated charge with blue and red color being level of positive and negative charges respectively. Therefore it is the hydrogen bond and hydrophobic forces which governs the interaction of 2-OG with Tpa1.

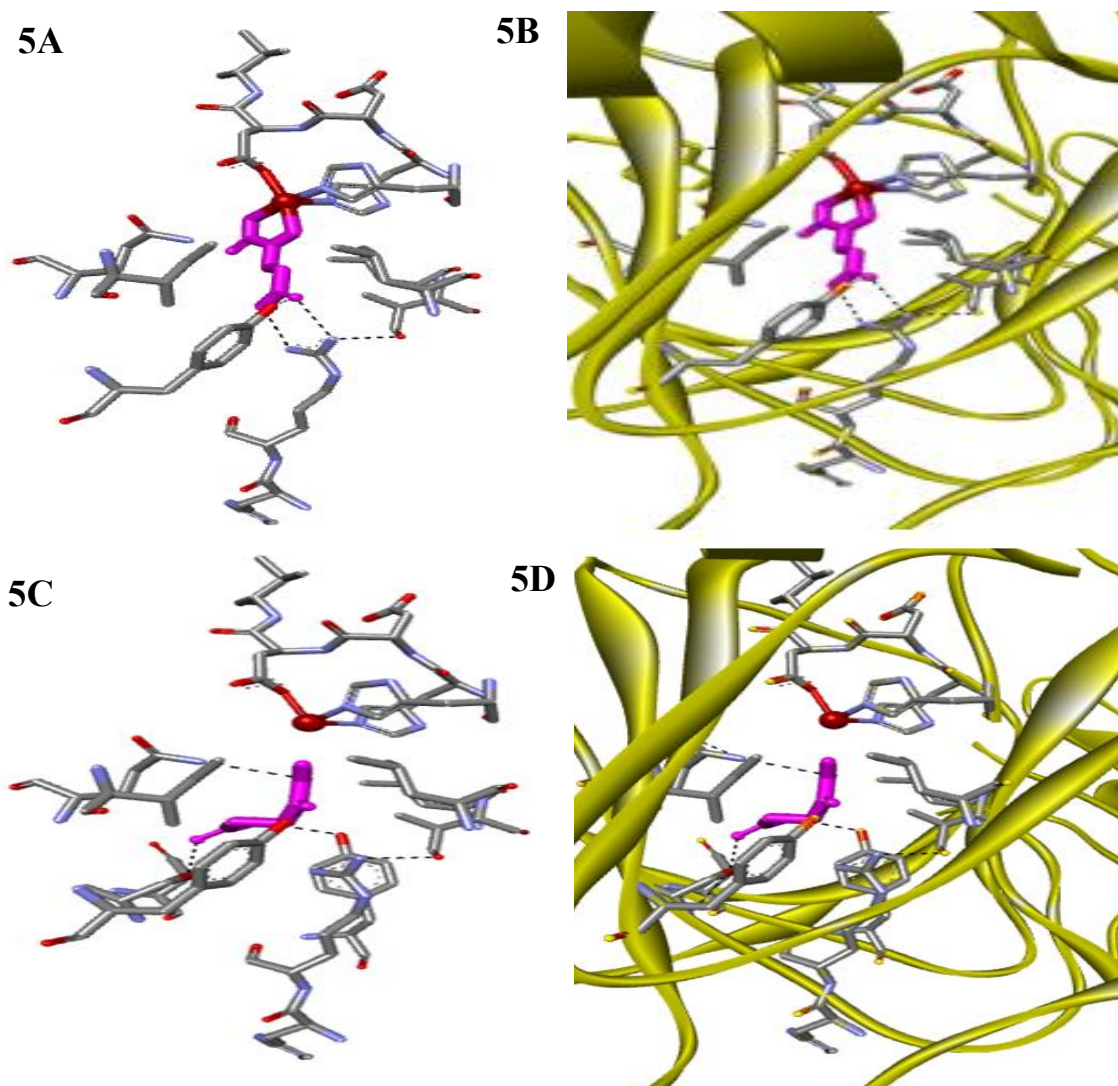
Molecular docking analysis and designing of functionally inactive Tpa1 mutations

We hypothesized that to generate functionally inactive Tpa1 mutant we have to design mutation in key Fe and 2OG binding residues such that the protein completely lose interaction with these two cofactors. To achieve this we decided to apply molecular docking analysis. Molecular docking analysis of WT Tpa1, Fe and 2OG was carried out using Autodock 4.2.2 and GLIDE docking suite from Schrodinger.

1.6.3 Molecular Docking WT Tpa1 with Autodock 4.2.2: The wild type Tpa1 protein was energy minimized using OPLS force field and the ligand (2-OG) was separated from the ternary complex. The separated 2-OG was then re-docked to the Tpa1 protein and the interactions were monitored. In order to validate the molecular docking, the docked ligand was superimposed to the native ligand and the r.m.s.d was calculated to be 1.98 Å. It was observed that the 2-OG interacts with Tyr150, Leu156, His159, Asp161, Ile171, Leu189, His227, Lys229, Arg238 and Ser240. Using Autodock 4.2.2 it was observed that 5 residues, including His159, Asp161, His 227, His 227 and Arg238 were consistently showing interactions for at least 10 different possible conformations of the ligand docked in the active site (**Figure 5**).

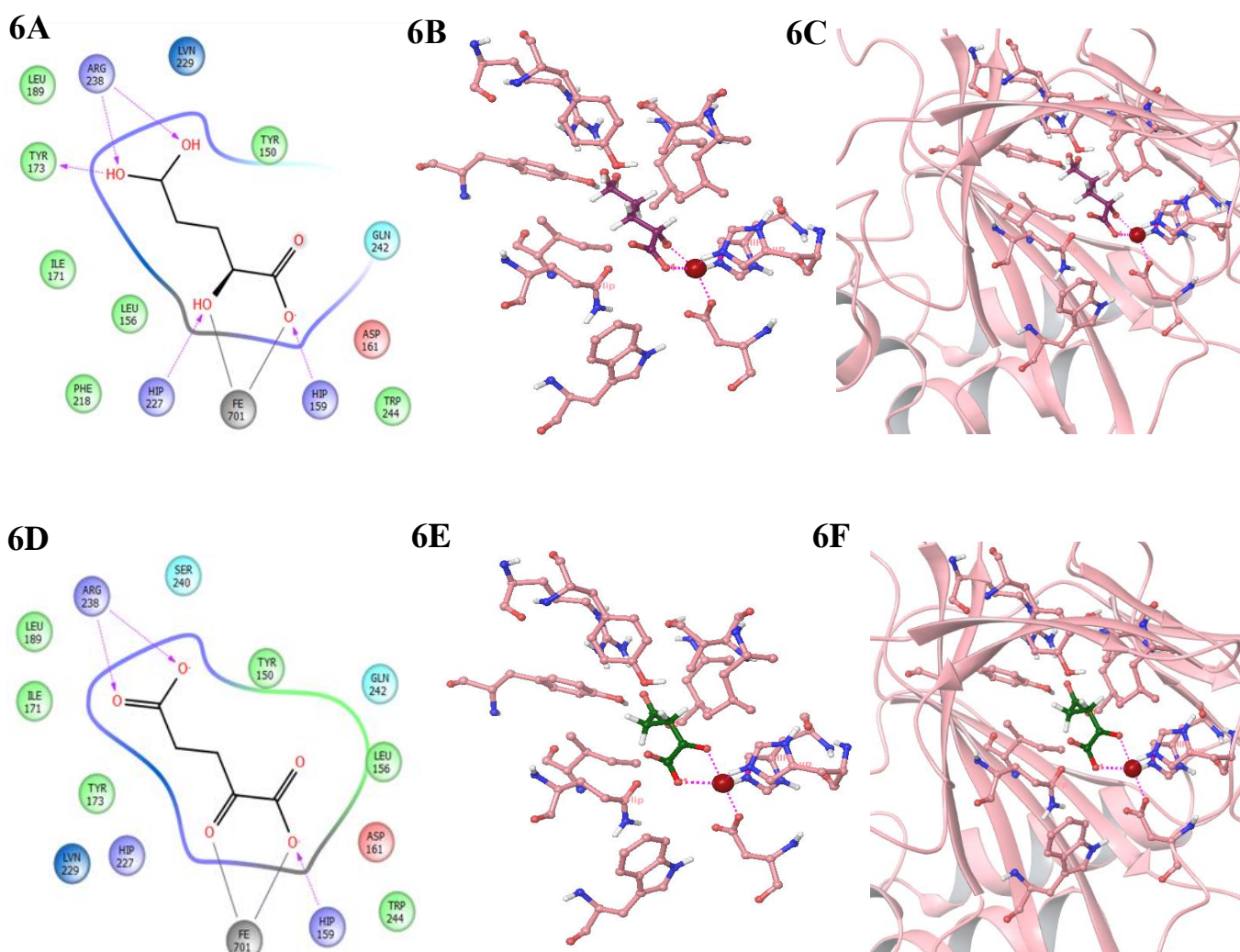
1.6.4 Molecular Docking of WT Tpa1 with GLIDE QPLD: Since the accuracy of docking the ligand with Autodock was less we considered docking using GLIDE QPLD (Quantum Polarized Ligand Docking). QPLD considers the polarization of the ligand before docking. The 2-OG with Quantum Mechanics modified charges using Jaguar was docked to the energy minimized Tpa1 protein (37). The docked poses were analyzed to study their interactions (**Figure 6**).

Figure 5



Molecular docking of Tpa1 using Autodock: The analysis of 2-OG interactions with Tpa1 crystal structure (5A and 5B). The ligand was removed from the ternary complex and redocked using Autodock to evaluate the reliability of the molecular docking algorithm (5C and 5D). It was observed that the residues Y150, R238, H237, L189, K229, Q242, Y173 and S240 shows interactions with the ligand. Amino acid residues of the Tpa1 protein are shown in grey and the secondary structure in yellow ribbons.

Figure 6



Molecular docking of Tpa1 with 2-OG using QPLD: Analysis of active site residues of wild type Tpa1 crystal structure (6A, 6B and 6C). The interactions of the ligand 2-OG whose charges were modified using Quantum Mechanics was redocked to the wild Tpa1 protein (6D, 6E and 6F). The docked ligand shows metal coordination with Fe similar to wild type crystal structure. The native ligand and docked ligand are represented in Magenta and green sticks respectively. 6A and 6D represents the 2D ligand interaction diagram from Schrodinger suite.

The docked 2-OG forms Hydrogen bonds with H159 and R238. The other residues namely, Y150, I171, Y173, L189, L156 and W244 were involved in hydrophobic interactions. While, S240 and Q242 makes polar contacts with 2-OG. The docking score was found to be -10.021 for the wild Ta1 protein.

1.6.5 Prediction of amino acid substitution by PROVEAN algorithm: To predict the amino acid substitution that would result in significant functional difference, the PROVEAN algorithm was employed. The impact of the amino acid variation is given by a delta alignment score, which is computed based on the BLOSUM62, amino acid substitution matrix. The lower the delta scores the more deleterious is the amino acid substitution. The default threshold of this algorithm is -2.5, the scores lesser than this affects the protein function drastically.

The output of the PROVEAN is mentioned below:

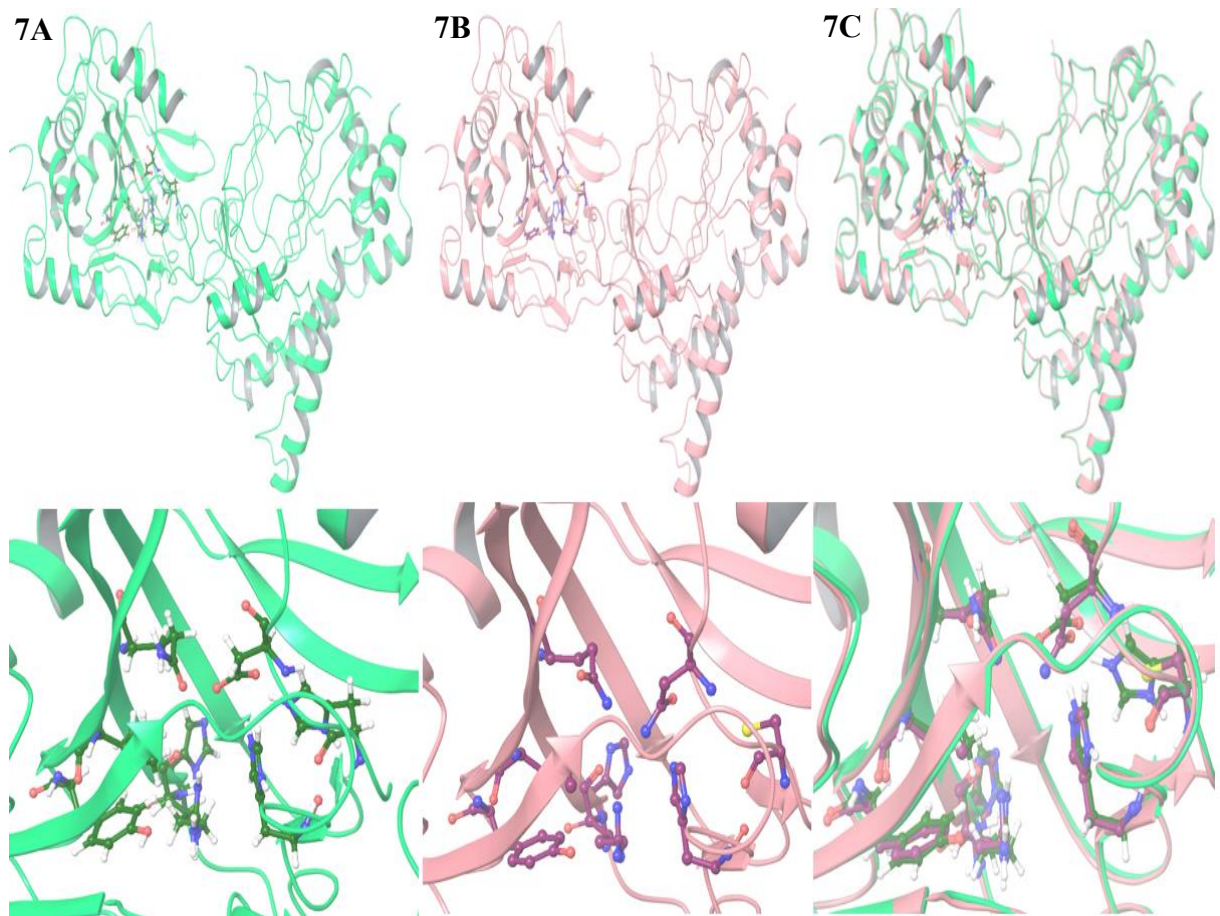
S.No	Substitution mutation	PROVEAN Scores
1	H159C	-11.00
2	D161N	-5.000
3	H227C	-11.00
4	H237C	-6.629
5	R238A	-5.913

1.6.6 *In silico* mutagenesis with PyMol: Based on the PROVEAN scores two Tpa1 mutant were generated: (a) H159C, D161N, mutant Tpa1 and (b) H159C, D161N, H227C, H237C, R238A mutant Tpa1. These substitution mutations were introduced in the Tpa1 protein using Site-directed mutagenesis wizard in PyMol. We observed that the r.m.s.d values of the H159C, D161N mutant (**Figure 7**) and H159C, D161N, H227C, H237C, R238A mutant (**Figure 8**) without the any bound cofactors were 4.317Å and 6.651Å, respectively and overall structure of these mutants were almost identical.

1.6.7 Molecular Docking of mutant Tpa1 with AutoDock:

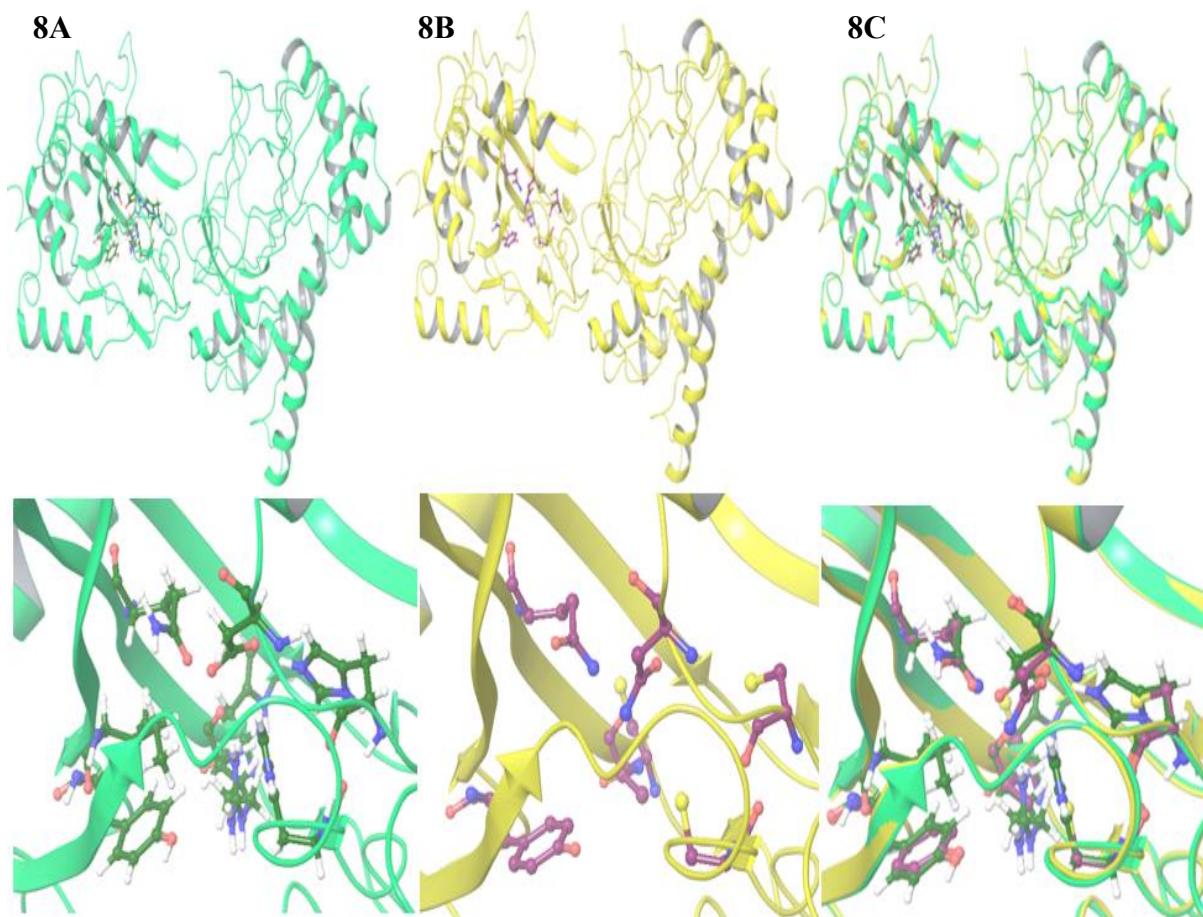
The prepared 2-OG was docked to the mutant tpa1 proteins namely, (a) H159C, D161N, mutant Tpa1 and (b) H159C, D161N, H227C, H237C, R238A mutant Tpa1 following the standard protocol involving flexible ligand and rigid protein. The grid box was as generated around the mutant tpa1 protein with dimensions same as the wild type Tpa1. Lamarckian genetic algorithm methodology was employed for docking simulations in AutoDock. The interaction pattern and thermodynamic stability of the docked ligand with mutant tpa1 was monitored. In H159C, D161N, mutant tpa1 the residues Y150, R238 and K229 forms hydrogen bonds with the 2-OG. Y173, S240, Q242 and W244 was showing other non-covalent interactions (**Figure 9**). Although the conformation of the ligand has changed, it was observed that the ligand was able to still interact with active site residues of the Tpa1 protein. Therefore we introduced 3 additional mutations in the active site to obtain completely abolish the Fe and 2-OG binding. Molecular docking analysis was then performed with H159C, D161N, H227C, H237C, R238A mutant Tpa1. It was observed that only few residues Y150, Y173, V229, L156 and Q241 was interacting with the 2-OG (**Figure 10**).

Figure 7



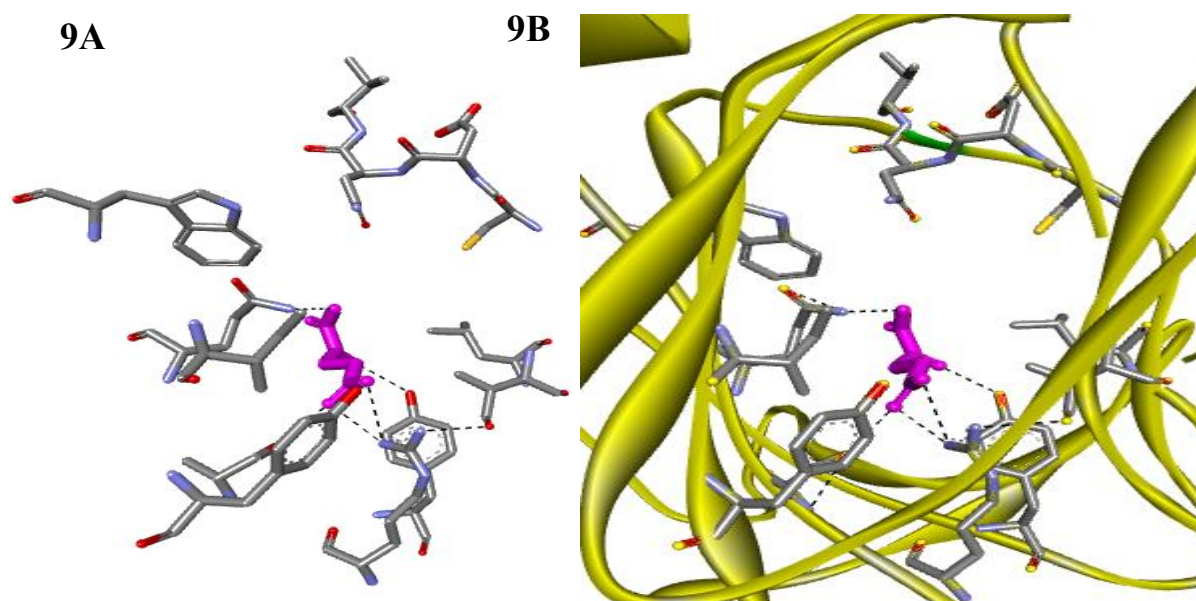
***In silico* mutagenesis using PyMol to generate Tpa1 with 2 mutations:** The residues interacting with the ligand were mutated using Site-Directed mutagenesis wizard in PyMol. Two Tpa1 mutant were generated namely H159C, D161N, mutant Tpa1. The wild type Tpa1 (7A) was superimposed on the mutant tpa1 (7B). The secondary structure of wild type Tpa1 and H159C, D161N, mutant Tpa1 are shown in green and pink ribbons respectively. The green and magenta sticks indicates the amino acid residues of wild and mutant Tpa1 respectively.

Figure 8



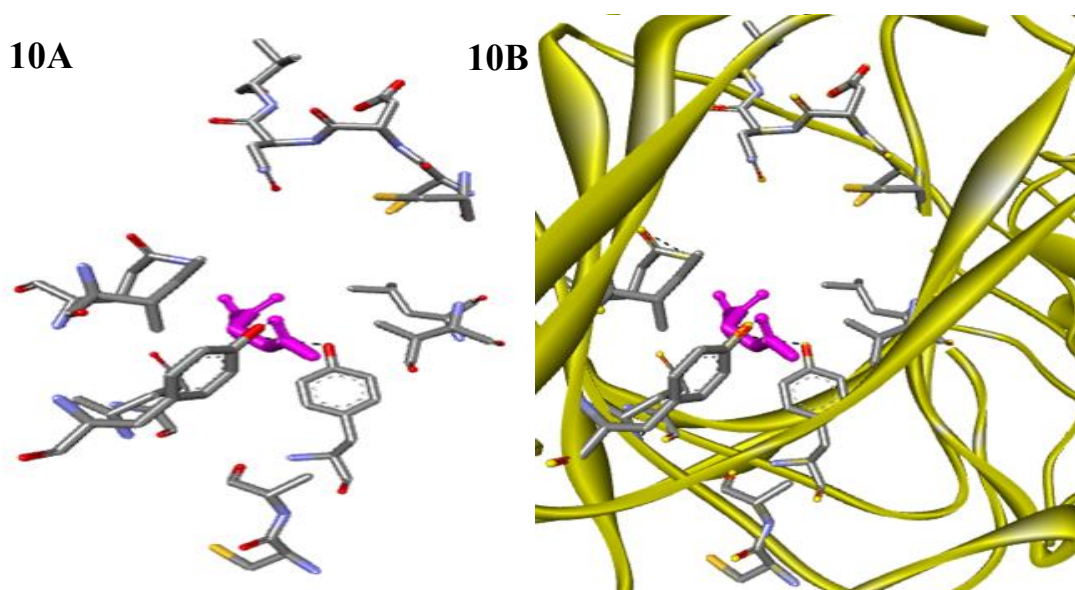
***In silico* mutagenesis using PyMol to generate Tpa1 with 5 mutations:** Site-Directed mutagenesis wizard in PyMol was used to mutate the active site residues in Tpa1 interacting with the ligand. The substitutions H159C, D161N, H227C, H237C, and R238A were incorporated in protein. The mutant generated (8B) was superimposed with wild type Tpa1 (8A), with the overall RMSD being 0Å. The secondary structure of wild type Tpa1 and H159C, D161N, mutant Tpa1 are shown in green and pink ribbons respectively. The green and magenta sticks indicates the amino acid residues of wild and mutant Tpa1 respectively.

Figure 9



Molecular docking of Tpa1 with 2 mutations using Autodock: Molecular docking analysis of 2-OG with mutant Tpa1 (H159C, D161N) structure (9A and 9B). The hydrogen atoms and Kollman charges were added to the optimized mutant protein structure. The prepared ligand was docked to the mutant protein using Lamarckian Genetic Algorithm. It was observed that the residues Y150, R238 and K229 were forming hydrogen bonds with the ligand. Residues Y173, S240, Q242 and W244 were showing other non-covalent interactions. It was also noted that the conformation of the ligand has changed and coordination with Fe is also lost. Amino acid residues of the Tpa1 protein are shown in grey and the secondary structure in yellow ribbons. The ribbons and sticks were represented using Discovery studio visualizer.

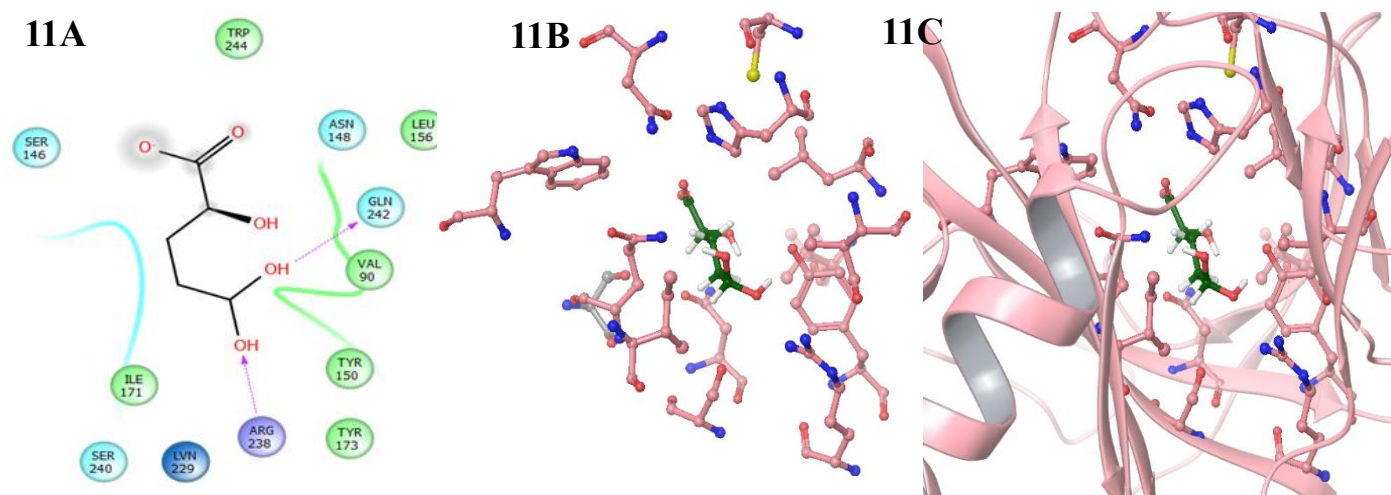
Figure 10



Molecular docking of Tpa1 with 5 mutations using Autodock: Analysis of 2-OG interactions docked with mutant Tpa1 (H159C, D161N, H227C, H237C and R238A) structure (10A and 10B) using Autodock. It was observed that the residues Y150, Y173, V229, L156 and Q241 was interacting with the ligand. Coordination of the ligand with Fe is abolished. The molecular docking analysis also shows that the conformation of the ligand has changed drastically. The interacting amino acid residues of the Tpa1 protein are shown in grey and the secondary structure in yellow ribbons using Discovery studio visualizer.

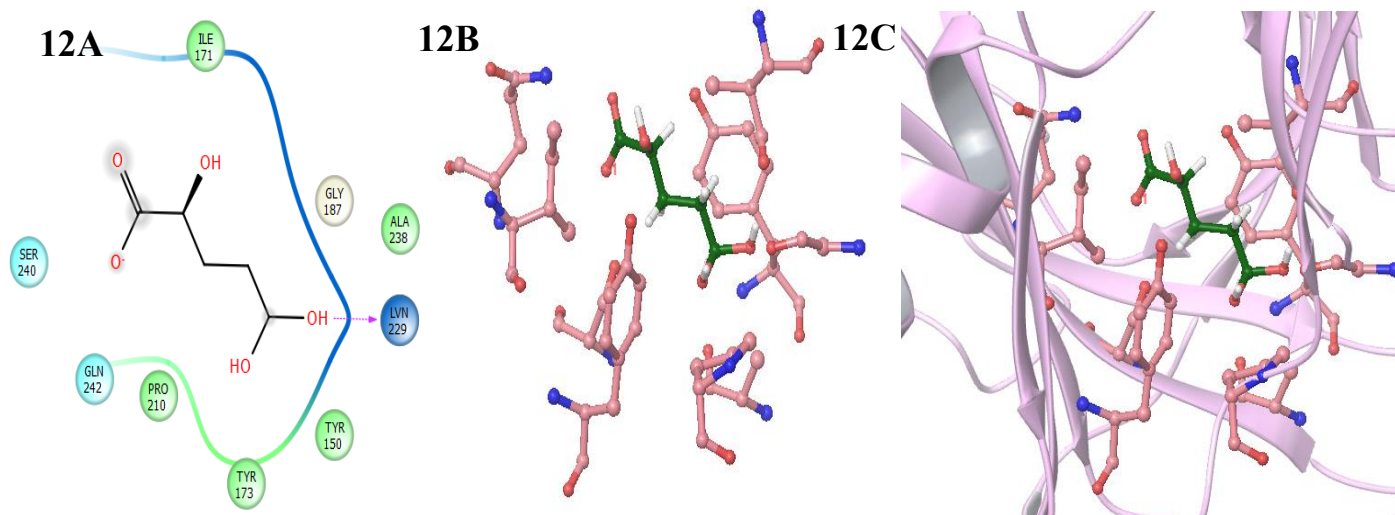
1.6.8 Molecular Docking of mutant Tpa1 with GLIDE QPLD: Since the accuracy of docking the ligand with Autodock was less, we therefore performed docking using GLIDE QPLD (Quantum Polarized Ligand Docking). This docking protocol considers the polarization of ligand. The QM/MM modified 2-OG was docked to the mutant tpa1 protein. The structure of the mutant Tpa1 proteins namely, H159C, D161N mutant Tpa1 and H159C, D161N, H227C, H237C, R238A mutant Tpa1 was optimized and energy minimized using OPLS_2005 force field. The overall potential energy of both the mutant protein after energy minimization was calculated to be -2666.558Kcal/mol. The molecular docking was first performed with Tpa1 protein containing 2 mutations (H159C, D161N). The analysis of the docked poses for 2 mutant showed that the ligand still forms hydrogen bond R238 similar to the wild type protein. Although the metal coordination with Fe was not observed, the 2-OG forms hydrophobic interactions with Y150, Y173, I171, L156, V90 and W244 (**Figure 11**). The analysis was performed with then performed with Tpa1 with 5 mutations. The coordination of 2-OG with Fe was completely lost and the hydrogen bonding pattern like the native Tpa1 protein was not observed. The mutant protein showed hydrogen bond with V229 only. It also showed hydrophobic interactions few residues namely, Y150, Y173, P210, A238 and P210 (**Figure 12**). The affinity of the 2-OG with Tpa1 mutants reduced drastically indicated by decrease in the docking scores for H159C, D161N mutant Tpa1 and H159C, D161N, H227C, H237C, R238A mutant Tpa1 being -4.7 and -3.95 respectively. Therefore, the functionally inactive Tpa1 could be designed by incorporating these substitution mutations in the active site of the protein.

Figure 11



Molecular docking of Tpa1 with 2 mutations using QPLD: The mutant Tpa1 (**H159C, D161N**) structure minimized using OPLS_2005 force field was docked with 2-OG using QPLD (11A, 11B and 11C). Although the metal coordination of ligand with Fe was not observed, residues R328 and Q242 still forms the hydrogen bonds similar to the wild type Tpa1. It was also noted that few residues namely Y150, Y173, S240, I171, V150, L156 and W244 interact with 2-OG. The conformation of the docked ligand has changed significantly compared to the wild type Tpa1. The pink and green sticks indicates the amino acid residues of mutant tpa1 protein and docked ligand respectively. 11A represents the 2D ligand interaction diagram from Schrodinger suite.

Figure 12



Molecular docking of Tpa1 with 5 mutations using QPLD: The energy minimized mutant Tpa1 (H159C, D161N, H227C, H237C, R238A) was docked with 2-OG using QPLD (12A, 12B and 12C). The docked 2-OG structure which was polarized using quantum mechanics (Jaguar fast from Schrodinger suite) showed an altered conformation than ligand in crystal structure. The docked poses showed very few interacting residues namely K229, Y150, Y173, P210, A238 and P210. The coordination of 2-OG with Fe was completely abolished. The pink and green sticks indicates the amino acid residues of mutant tpa1 protein and docked ligand respectively. 12A represents the 2D ligand interaction diagram from Schrodinger suite.

PART-II

2. Protein Docking and Prediction of Molecular Interactions of *E. coli* AlkB

2.1 Introduction

Alkylating agents from the environment such as tobacco specific nitrosamines [39], products of cellular metabolism like free radicals generated during lipid peroxidation and chemotherapeutic drugs used in cancer treatment such as Temozoloamide, Melphalan causes alkylation damage in the DNA. Alkylating agents can react with 12 different sites of the bases as well as with phosphate backbone of DNA generating methyl phosphoesters (MTE) . To overcome this damage cells have evolved genes that forms proteins that are involved in the repair of with alkylation specific DNA damage. In *Escherichia coli*, resistance to alkylation damage is because of increased expression four genes namely, *ada*, *alkB*, *alkA* and *aidB*. This response of cells to the alkylation damage is called the adaptive response [40]. Expression of these genes is mainly regulated by *Ada*, and their induction provides protection against alkylation damage to DNA. Cysteine residues, namely Cys-38 and Cys-321 of the *Ada* protein are the important residues which mediates the demethylation of DNA [41]. O6 -methylguanine and O4-methylthymine lesions are removed by Cys-321, which acts as the methyl acceptor [42]. Cys-38 is required for demethylation of phosphomethyltriesters (PME) in the sugar phosphate backbone of the DNA. Transfer of the methyl group to the Cys-38 residue converts *Ada* into a transcriptional activator and hence controls the expression of the *ada* regulon. The *ada alkB* operon is involved in the transcription of both *ada* and *alkb* genes which is under the control of the *ada* promoter. *AlkB* mutants

are sensitive to the alkylating agents such as methyl methanesulfonate (MMS) and dimethyl sulfate. The *alkA* gene encodes AlkA, which is a glycosylase involved in the repair of lesions including N7 -methylguanine and N3 -methyl purines and O2 -methyl pyrimidines. The function of the *aidB* gene is still unknown but appears to be involved in the inactivation of endogenous nitrosamines.

2.1.1 Role of AlkB in the repair of damaged ssDNA

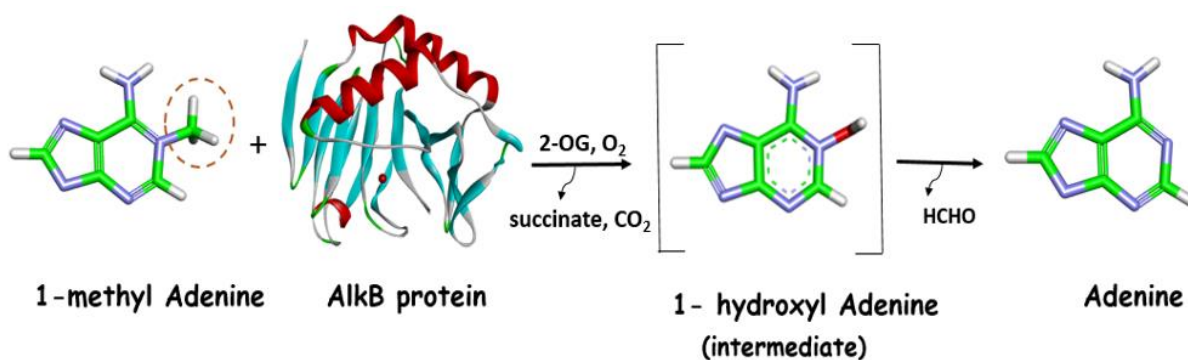
Escherichia Coli Alkylation protein B (AlkB) is a non heme iron-dependent oxidoreductase and member of 2-oxoglutarate (2OG)-dependent dioxygenases [43]. In human there are 9 homologs of AlkB namely ALKBH 1-8 and FTO, in which ABH1, ABH2 and ABH3 have DNA repair activity. AlkB is an oxidative demethylase which is involved in the direct repair of cytotoxic lesions namely 1-methyladenine and 3-methylcytosine as well as the exocyclic adducts such as 1, N6-ethenoadenine (ϵ A) [44]. AlkB operates by oxidative demethylation reaction which requires Fe (II) and α -ketoglutarate as cofactor and co-substrate respectively. It catalyzes the decarboxylation of α -ketoglutarate to form succinate and CO₂ coupled with demethylation of the damaged base thereby restoring the undamaged adenine and cytosine residues in the DNA.

2.1.2 Mechanism of repair of damaged base by AlkB protein

In the catalytic core of AlkB protein Fe ion is initially coordinated with three water molecules thereby stimulating the binding of α -ketoglutarate to the active site. This binding of α -KG displaces two water molecules to create the Fe (II)/ α -KG active site

complex. In the presence of dioxygen the third water molecule is displaced, resulting in the nucleophilic attack on the α -keto group. This results in decarboxylation of α -KG producing succinate and CO_2 coupled with hydroxylation of the methylene group resulting in a hydroxymethyl intermediate thereby restoring the undamaged nucleobase.

a)



b)

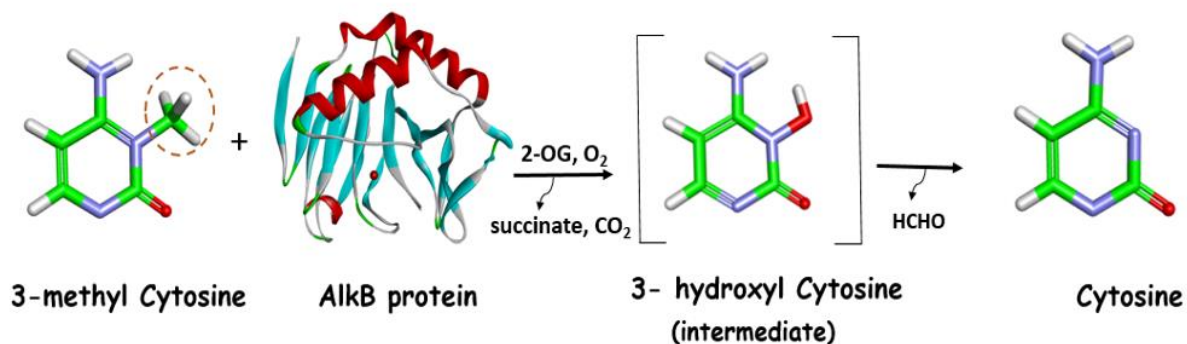


Figure1: Mechanism of oxidative demethylation of 1-meA (a) and 3-meA (b) by DNA repair protein AlkB. In the presence of 2-OG, Fe (II) and dioxygen, AlkB catalyzes the removal of methyl group from the bases thereby releasing succinate and CO_2 as

byproducts. The bases are represented in sticks with carbon, nitrogen and oxygen colored coded with green, blue and red respectively.

2.1.3 Structure of AlkB protein

E.coli AlkB is a 216 amino acids protein which has a characteristic Double Stranded Beta Helix (DSBH) protein fold that involves eight beta strands forming a “jelly roll” [45]. The N-terminal region of AlkB from 1-90 amino acids constitutes the nucleotide recognition lid which in the absence of DNA is flexible to access the substrate. The damaged base is bound in the hydrophobic cavity where it is held by the base stacking interaction with His131 and Trp69 from the nucleotide recognition lid. In the absence of bound substrate residue Y76 is flipped away from active site thereby allowing the damaged base to access the catalytic core. The T51-Y55 loop shifts about 2.5Å from the active site upon binding of the substrate. In the closed conformation the Y76 residue interacts with the damaged base and clamps it into the positions for the catalysis [46].

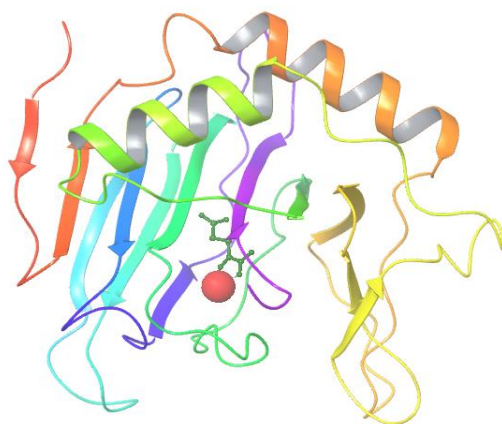


Figure2: Crystal structure of AlkB protein (PDB ID: 1QVC) at 2.20 Å resolution. The characteristic DSBH fold with eight beta sheets running in the antiparallel direction. The Fe (II) ion and 2-OG are represented in red sphere and green sticks respectively.

AlkB has the characteristic H₁₃₁ D₁₃₃...H₁₈₇ motif which is important for the coordination of Fe atom in the catalytic core. In addition to the characteristic HDH triad, AlkB also has the RXXXXXXR motif where the first arginine residue is involved in the coordination of Fe ion while the other arginine is involved in AlkB-specific substrate binding [47]. The carboxylate moiety alpha-ketoglutarate forms salt bridge with Arg204 and Arg210 residues of AlkB protein.

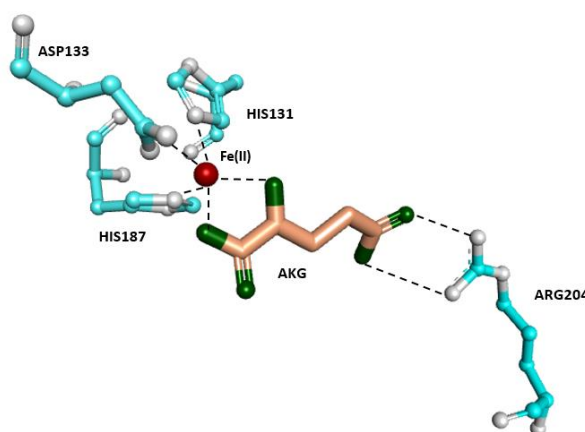


Figure3: The catalytic core of AlkB protein illustrating the characteristic HDH motif and coordination of Alpha-ketoglutarate with Fe ion. AlkB and Fe ion are shown in cyan sticks and red spheres respectively.

AlkB preferentially binds and repairs the damaged ssDNA compared to the dsDNA, the reason behind which is still an open question. Also, in vivo the ssDNA are always bound majorly by SSB and RecA proteins. Therefore, the objective is to understand how the SSB and RecA coated ssDNA could be a substrate for AlkB thereby enhancing its repair activity. This question was thus addressed using the *Insilco* docking programs namely PatchDock, ClusPro and ZDOCK.

2.1.4 Role of Single-stranded DNA-binding protein

Bacterial Single-stranded DNA-binding (SSB) proteins are a class of nonspecific DNA binding proteins. SSB is also called as “helix-destabilizing proteins” plays an important role in the DNA replication, repair and recombination. During replication of the DNA, SSB protein stabilizes replication origins, it prevents reannealing and formation secondary structures by the single stranded DNA, it protects against single-strand nuclease digestion, it aids helicases in helix destabilization, it helps to organize and it is also required for the primosome assembly [48]. In vitro methyl directed mismatch repair has an absolute requirement for SSB. The three key enzymes in mismatch repair namely; DNA helicase II, exonuclease I and DNA polymerase III holoenzyme interact functionally with SSB. In vitro recombination involves 3 main steps namely presynapsis, synapsis and strand exchange. During heteroduplex formation, SSB binds to the displaced strand thereby preventing reannealing.

2.1.5 Structure of Single Strand DNA Binding (SSB) protein

Each monomer of E.coli SSB consists of 176 amino acids and has the characteristic OB (Oligosaccharide Binding fold) [49]. SSB protein is composed of characteristically two domains namely, an N-terminal domain which is involved in the binding of ssDNA and a C-terminal domain. The CTD of SSB is enriched in glycine and acidic amino-acids. A short hexapeptide motif with sequence D-D-D-I/L-P-F is present near the end of the protein [50]. The core of the protein monomer structure is hydrophobic in nature.

The N-terminal region from residues 1-105 are the highly ordered regions of SSB protein. The region from 106-176 forms a long random coil which is mainly constituted by glycine, proline, glutamine and asparagine residues.

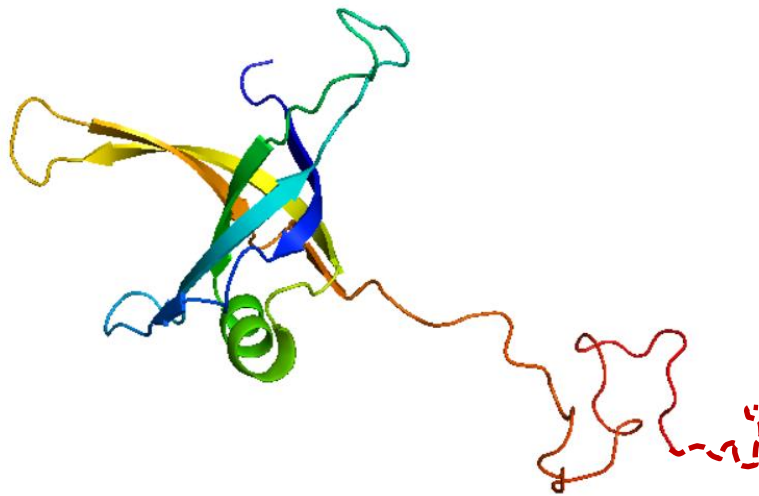


Figure 4: Crystal structure of monomeric Single Strand DNA Binding (SSB) protein (PDB ID: 1QVC) at 2.20 Å resolution. The broken lines indicates the residues (146-176) that could not be crystallized because of poor electron density clouds.

The quaternary structure of SSB is formed in two different ways depending upon the interface regions namely L45 loop-mediated tetramer and beta-strand mediated. In case of L45 loop-mediated tetramer, the L45 constitutes the interface of the tetramer while in the beta-strand mediated tetramer the six beta strand constitutes the packing interface of the tetramer. Although two different

arrangements of the SSB tetramer is possible in the crystal structure, in solution the six-stranded β -sheet-mediated tetramer configuration exists [51].

Four monomers of SSB associates to form a stable homotetramer that has 2 different binding modes which is denoted as $(SSB)_n$ where “n” represents the number of nucleotides occluded per tetramer. In SSB_{35} binding mode, only two monomers of SSB interacts with ssDNA, while in $(SSB)_{65}$ binding mode, all the four monomers interacts with ssDNA [52].

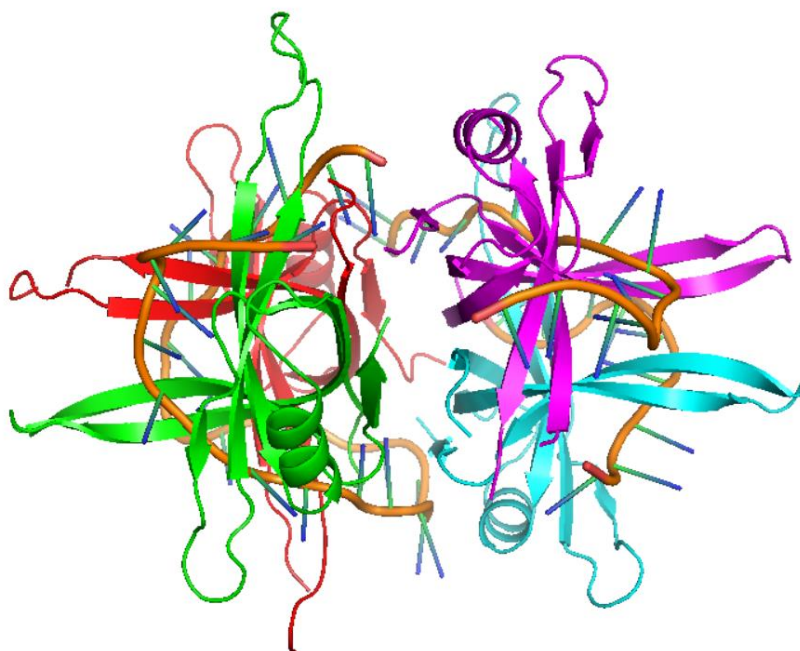


Figure 5: Crystal structure of SSB tetramer wrapped with ssDNA (PDB ID: 1EYG) at 2.80 Å resolution. This is six-stranded β -sheet-mediated tetramer with each monomer represented in blue, green, red and magenta respectively. The DNA backbone is represented in orange color.

2.1.6 Structure of RecA protein

RecA monomer is 352 amino acids protein which folds into three characteristic domains namely, the N-Terminal Domain (NTD), the core domain and the C-Terminal Domain (CTD) [52]. The NTD residues 1–33 comprises of a single alpha helix and a beta strand which is involved in the formation of RecA polymer. The core residues (34-294) consisting of a central beta sheet with 6 alpha helices is involved in ATP binding. The loops L1 and L2 of the core domain are involved in the binding of ssDNA. The CTD of RecA is involved in the interfilament interactions and consists of 3 alpha helices and 3 beta sheets [53].

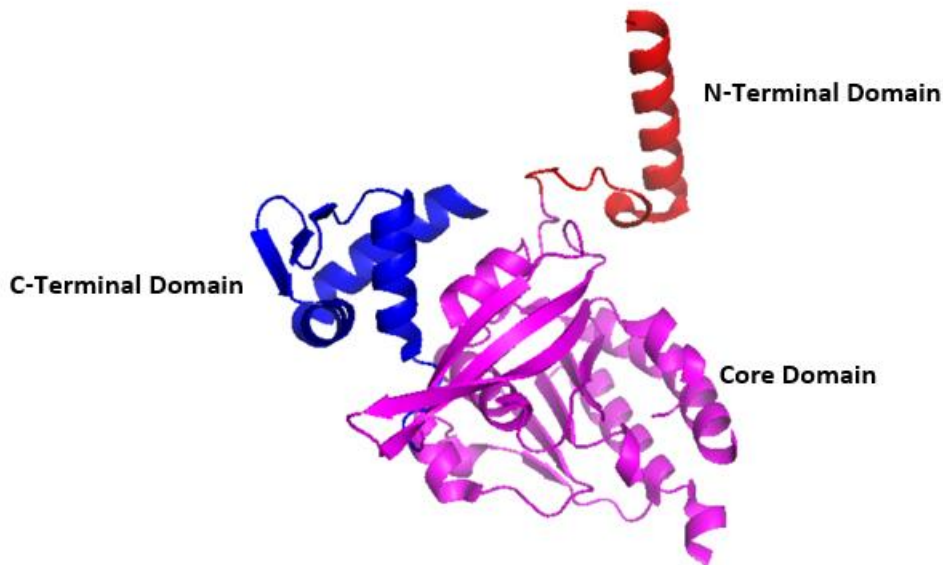


Figure 6: Crystal structure of RecA monomer (PDB ID: 2REB) at 2.30 Å. The NTD (6–33), core (34-294) and CTD (270–328) are represented in red, magenta and blue cartoon respectively.

ATP bound RecA polymerizes on ssDNA to form a right-handed helical filament which is formed such that each protomer binds three nucleotides of ssDNA. The RecA bound DNA is stretched with an average rise of 5.1 Å per nucleotide [54].

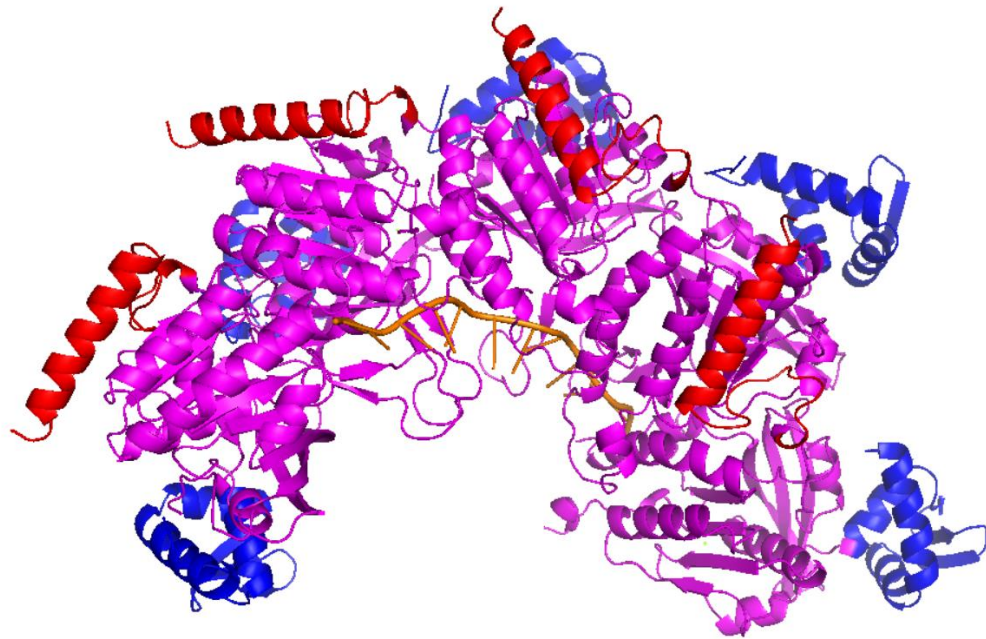


Figure 7: Crystal structure of RecA nucleoprotein filament (PDB ID: 3CMW) at 2.80 Å resolution. The NTD (6–33), core (34–294) and CTD (270–328) are represented in red, magenta and blue cartoon respectively. The RecA bound ssDNA which is extended is represented in orange color cartoon.

2.2 Methods and Materials:

2.2.1 Protein Structure Preparation

The atomic coordinates of the protein files namely AlkB (3KHC), RecA (3CMW) and SSB (1EYG) were retrieved from Protein Data Bank. The Protein Preparation Wizard (PrepWizard) of the Schrodinger suite (Schrödinger, LLC, New York, NY, 2014) was implemented to preprocess, which assigns the bond orders and adds hydrogen to the protein structure. The protein was optimized using PROPKA, which predicts the pKa of the ligand for accurate molecular docking (9). The optimized protein was then energy minimized using Optimized Potentials for Liquid Simulation (OPLS_2005, Schrödinger, New York, NY, and USA) force field.

2.2.2 Docking analysis using PATCHDOCK

The AlkB protein was docked into RecA filament using Patchdock online server which employs the algorithm that generates the output complexes based on the shape complementarity of the biomolecules [55]. For this, the 3D structure of SSB and RecA coated DNA receptor files were submitted to Patchdock along with AlkB. The top 20 complexes generated were obtained based on the 2 parameters namely atomic contact energy (ACE) and geometric shape complementarity score. The docked poses were further subjected to FireDock for the post energy minimization.

2.2.3 Docking analysis using ZDOCK

ZDOCK, a rigid-body protein–protein docking tool was employed to dock SSB ssDNA complex and RecA ssDNA complex with AlkB protein. It uses the Fast Fourier Transform (FFT) algorithm to perform the global docking analysis. This docking program involves a combination of shape complementarity, electrostatics and statistical potential terms for scoring of the docked poses [56].

2.2.4 Docking analysis using CLUSPRO

ClusPro is a web-based program for docking the protein structures. The structure of the proteins to be docked is submitted through ClusPro's web interface. Cluspro is a fully automated rigid-body docking tool which filters the docked conformations based on the clustering properties. This involves evaluation of empirical free energy to select the docked pose with lowest desolvation and electrostatic energies [57]. First, the docking algorithms evaluates structures with favorable surface complementarities. Then, structures with good electrostatic and desolvation free energies for further subjected to further clustering.

2.2.5 Analysis and visualization of docking simulation results

Docking analysis was performed to identify the possible conformations and orientations of the AlkB protein around SSB ssDNA and RecA ssDNA complex. The best conformation of the AlkB around the protein-DNA complex was selected based on two criteria: shape complementarity and electrostatic binding affinity of the complexes. The best poses were analyzed for non-covalent interactions namely hydrogen bonding, Vander Waals interactions and hydrophobic interactions using Discovery Studio

Visualizer 2.5 and PyMol. The inter atomic distances between the two interacting proteins were monitored using WhatIf server.

2.3 Objective

1. To study whether the RecA bound DNA is a substrate for AlkB using *insilico* molecular docking analysis.
2. To understand how SSB bound DNA could be a substrate for AlkB using *insilico* molecular docking programs.

2.4 Results and Discussion

2.4.1 Docking of AlkB with ssDNA wrapped SSB tetramer

The optimized structure of AlkB protein was docked with ssDNA bound RecA filament using 3 docking tools namely PatchDock, ZDOCK and ClusPro. The top clustering outputs from each of these programs were considered for further analysis. Docking of ssDNA wrapped SSB tetramer with AlkB using PatchDock gave the Score and ACE values of 14514 and -167.23 respectively. Global Energy value and Attractive Vander Waals energy using FireDock is -17.33Kcal/mol and -22.86Kcal/mol respectively. Analysis with ClusPro docking tool gave the clustering scores of -755.4 with SSB-AlkB complexes. The Docking analysis shows that the bases in the ssDNA are in close proximity with nucleotide recognition lid of the AlkB protein. The non-covalent interactions of the RecA and SSB coated ssDNA with AlkB was monitored using Protein-Protein interaction analysis tool from Schrodinger. The interatomic distances between the bases and the AlkB protein was studied using the WhatIf server. It was observed that the residues of AlkB Tyr76, Leu130, Arg161, Lys127 and Tyr186 interacts with Met23, Arg21, Glu65, Glu69 and Tyr22 residues of SSB protein. It was observed that bases from polyC ssDNA namely Cyt10, Cyt11, Cyt15 and Cyt28 interacts with gln74, Pro52, Arg140 and Lys134 of AlkB residues respectively. It was also noted that the bases from ssDNA wrapped SSB tetramer are in close proximity with AlkB protein. The AlkB protein adopts an energetically favorable conformation and interacts without any steric hindrance. Since the docking methodology employed involves the rigid protein-protein docking, the flexibility of the loop regions of the AlkB molecules were not considered. Thus, we assume that AlkB is recruited through the

protein-protein interaction with SSB and RecA which provides a stable platform for AlkB to interact with ssDNA thereby enhancing its DNA repair activity.

Docking of AlkB with SSB tetramer wrapped with ssDNA using ClusPro

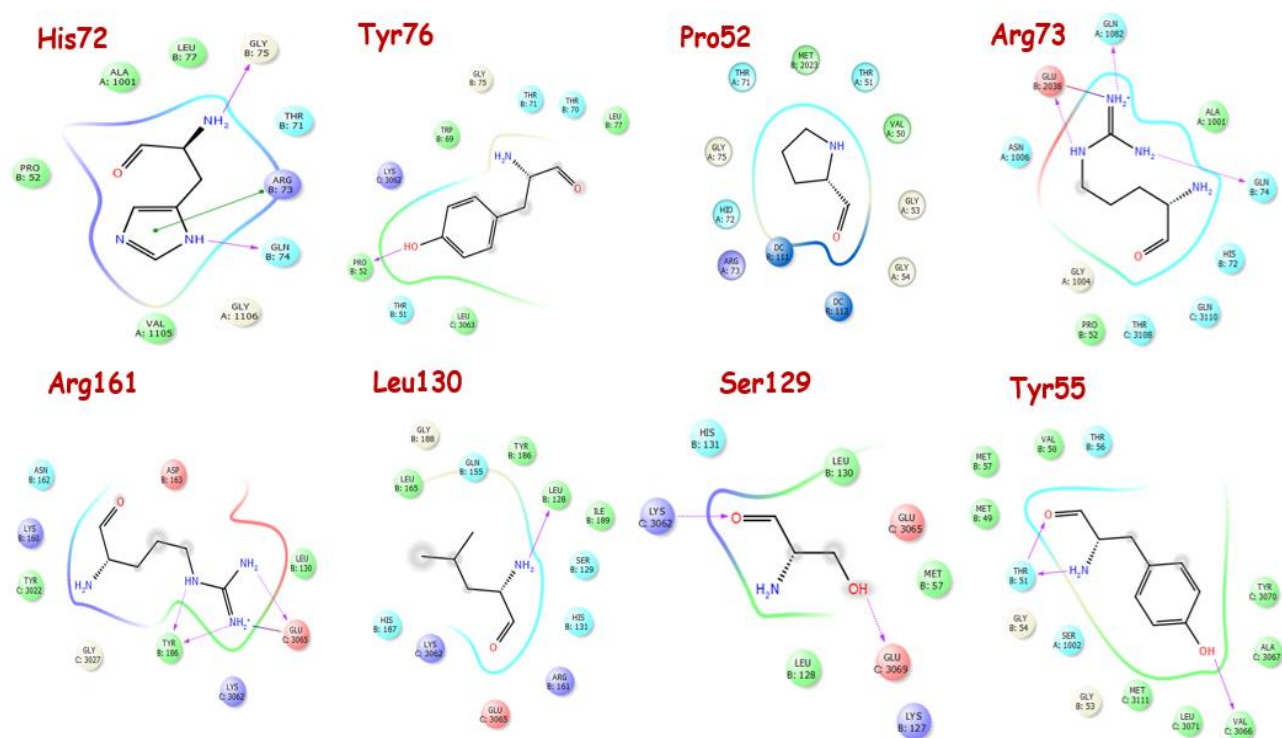


Figure 8: Docked pose of AlkB with ssDNA wrapped SSB tetramer generated using ClusPro. SSB tetramer, AlkB and ssDNA are represented in magenta, cyan and orange colors respectively.

Tabular column 1: AlkB residues interacting with ssDNA wrapped SSB tetramer

S.No	AlkB residues	SSB residues	Chain	Distance Å	Nature of interaction
1	His72	Val105	B	3.56	Hydrophobic
2	Arg73	Glu38 Gln82	B A	3.27 -	Salt bridge Hydrogen bonding
3	Thr70	Trp40	B	3.32	Hydrophobic Electrostatic
4	Gln84	Trp40	B	3.27	Hydrophobic
5	Lys134	Ala38	C	3.54	Electrostatic
6	Tyr76	Lys62 Leu63	C C	3.73 3.43	Electrostatic Hydrophobic
7	Leu130	Glu65	C	3.32	Hydrogen bonding
8	Arg161	Glu65	C	3.23	Hydrogen bonding
9	Tyr55	Val66	C	3.41	Electrostatic
10	Lys127	Glu69	C	2.81	Hydrophobic

Figure 9: 2D interaction diagram of AlkB with ssDNA wrapped SSB tetramer



Docking of AlkB with SSB tetramer wrapped with ssDNA using PatchDock

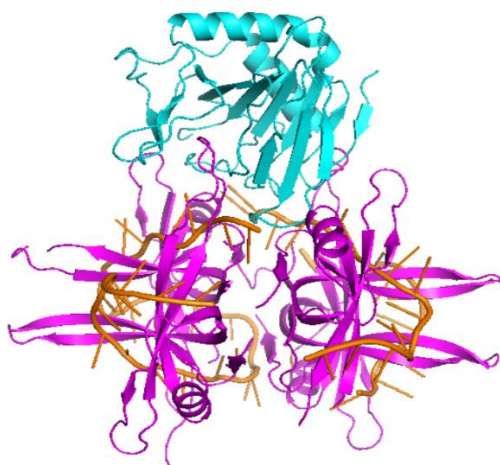


Figure 10: Docked pose of AlkB with ssDNA wrapped SSB tetramer generated using PatchDock. SSB tetramer, AlkB and ssDNA are represented in magenta, cyan and orange colors respectively.

Figure 11: 2D interaction diagram of AlkB with SSB tetramer

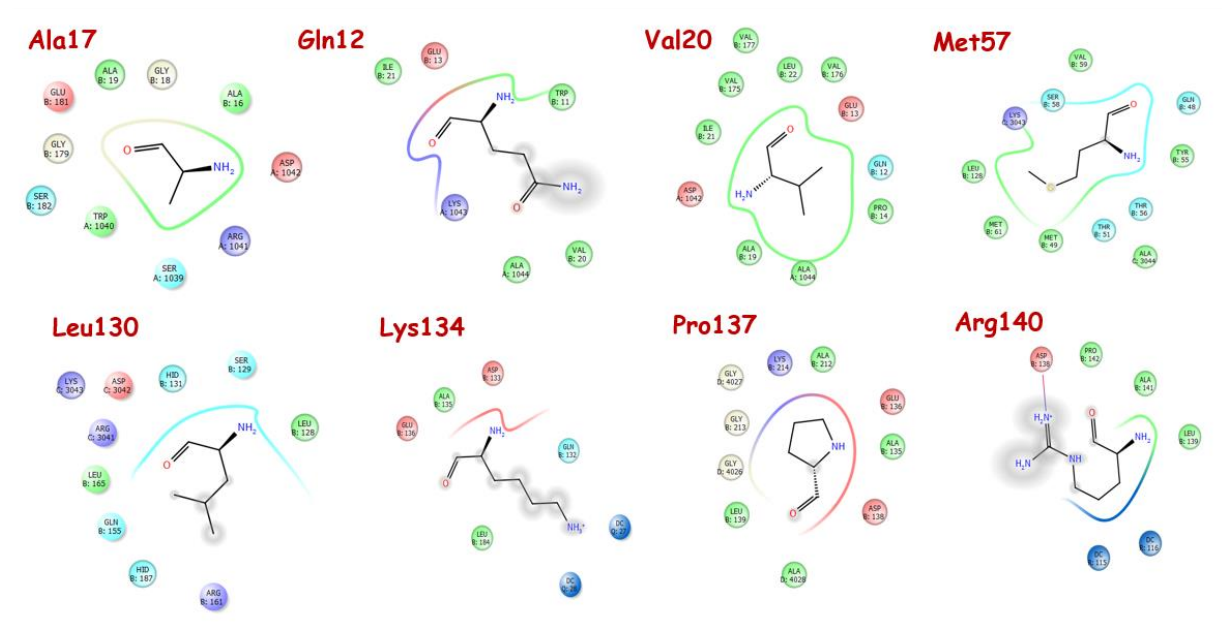
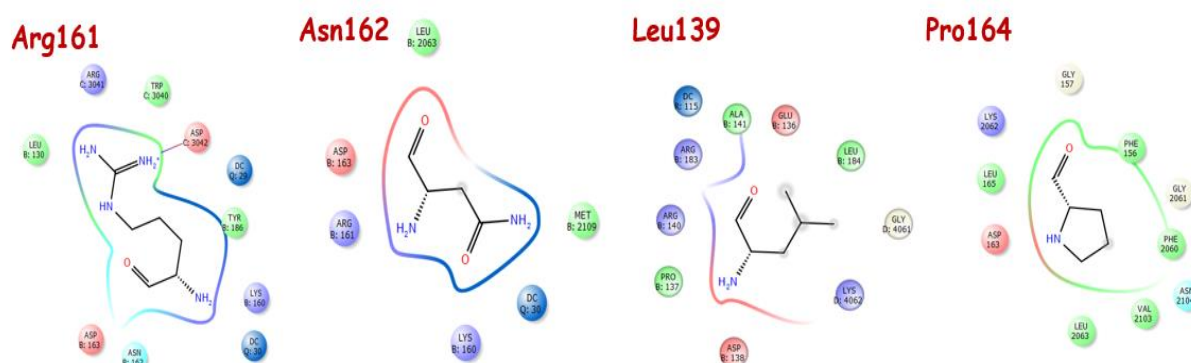


Figure 12: 2D interaction diagram of AlkB with SSB tetramer



Tabular column 1: Residues of AlkB interacting with SSB tetramer

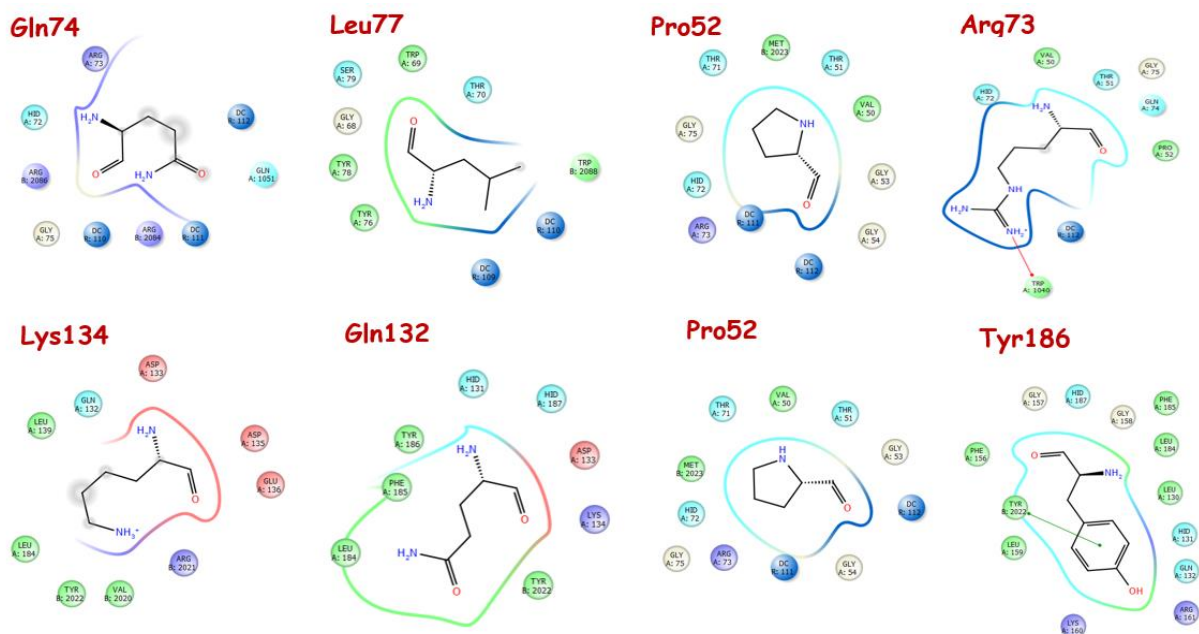
S.No	AlkB residues	SSB residues	Chain	Distance Å	Nature of interaction
1	Ala17	Trp40	A	3.16	Hydrophobic
2	Gln12	Lys43 Ala44	A A	3.54 3.45	Electrostatic Hydrophobic
3	Val20	Ala44 Asp42	A A	3.37 3.45	Hydrophobic Electrostatic
4	Met57	Ala44	A	3.75	Hydrophobic
5	Leu130	Arg41	A	3.32	Electrostatic
6	Lys134	Cyt27 Cyt28	Q Q	3.3 3.64	Electrostatic Electrostatic
7	Pro137	Gly26 Ala28	D D	2.93 3.5	Hydrophobic
8	Leu139	Lys62	D	2.34	Electrostatic
9	Arg140	Cyt15	R	2.99	Electrostatic
10	Arg161	Trp40	C	3.62	Hydrophobic

Docking of AlkB with SSB tetramer wrapped with ssDNA using ZDock

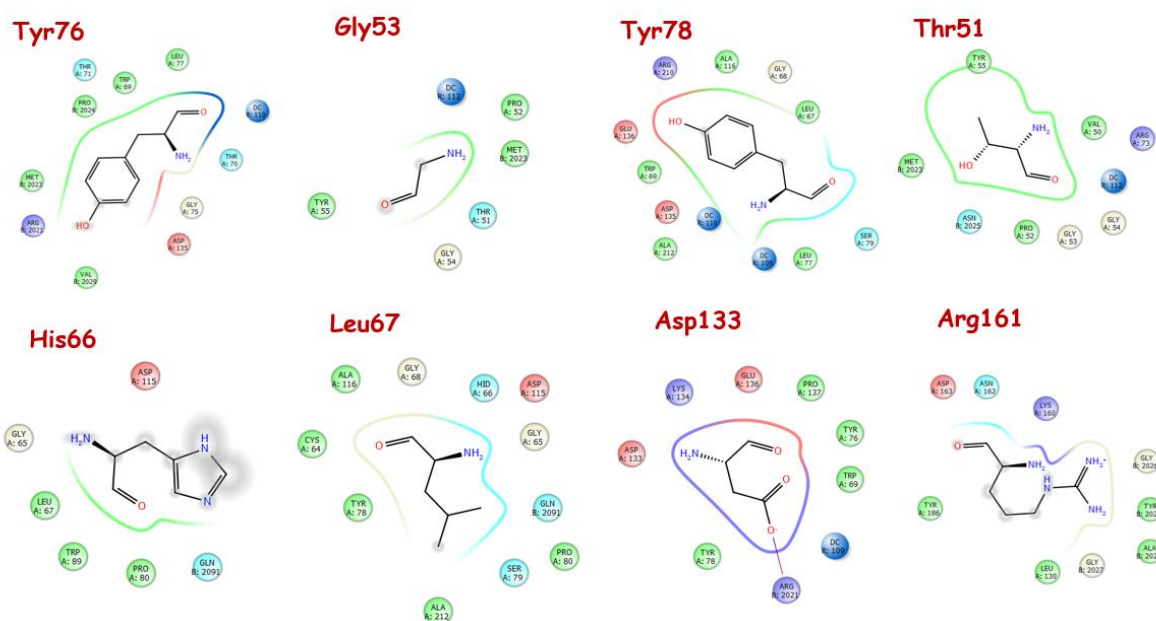


Figure 12: Docked pose of AlkB with ssDNA wrapped SSB tetramer generated using PatchDock. SSB tetramer, AlkB and ssDNA are represented in magenta, cyan and orange colors respectively.

Figure 13: 2D interaction diagram of AlkB residues with SSB tetramer



2D interaction diagram of AlkB residues interacting with SSB tetramer



Tabular column 1: Residues of AlkB interacting with ssDNA wrapped SSB

tetramer

S.No	AlkB residues	SSB residues	Chain	Distance Å	Nature of interaction
1	Lys134	Val20 Arg21	B B	3.07 3.35	Hydrophobic Electrostatic
2	Tyr76	Cyt11	R	3.79	Electrostatic
3	Asp135	Arg21	B	2.99	Salt bridge
4	Pro52	Met23 Cyt11, Cyt12	B R,R	3.40	Hydrophobic Electrostatic
5	Gln74	Arg84 Cyt10	B R	3.30 3.43	Electrostatic Electrostatic
6	Leu77	Trp88 Cyt9	B R	3.45 3.11	Hydrophobic Electrostatic
7	Arg73	Trp40 Cyt12	B R	3.45 3.61	Pi-cation interaction Hydrophobic
8	His66	Gln91	B	3.34	Polar contact
9	Tyr186	Tyr22	B	3.63	Pi-pi stacking interaction
10	Tyr78	Cyt9	B	3.18	Electrostatic

Docking of AlkB with ssDNA wrapped SSB tetramer

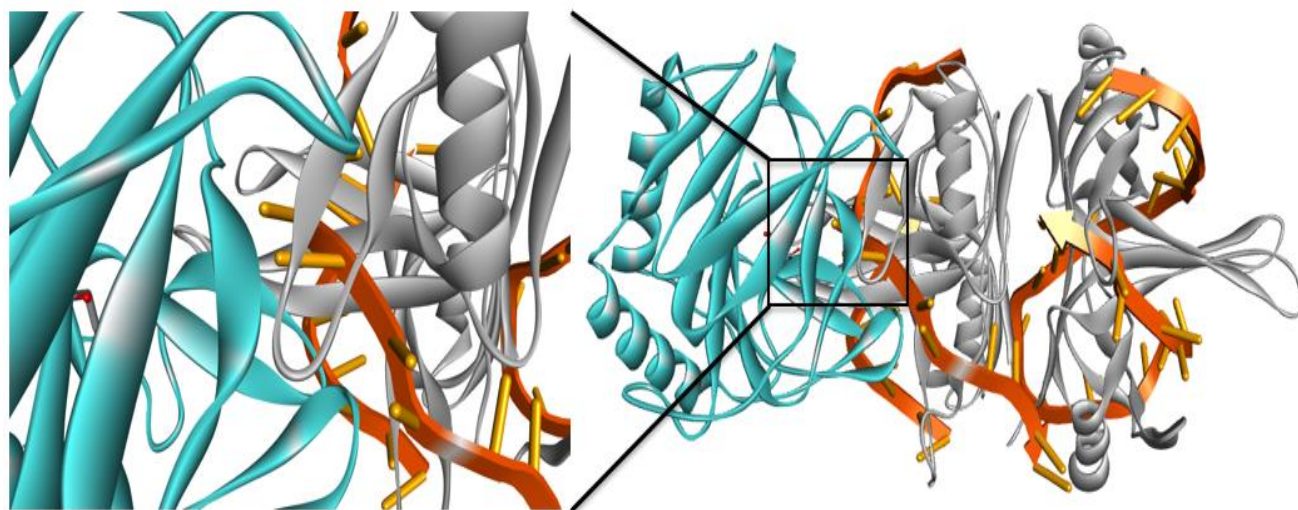


Figure 14: Model illustrating the interaction of AlkB with ssDNA wrapped SSB tetramer. *In silico* docking analysis using ZDOCK shows that AlkB could dock in close proximity with ssDNA without any steric hindrance and adopt an energetically favorable conformation. The AlkB protein and SSB tetramer are represented in cyan and grey color cartoon respectively. The sugar phosphate backbone of ssDNA are represented in orange color.

2.4.2 Docking of AlkB with ssDNA bound RecA filament

AlkB protein which was optimized using Schrodinger Protein-preparation wizard was docked with structurally optimized ssDNA bound RecA filament using 3 docking tools namely PatchDock, ZDOCK and ClusPro. The top clustering outputs from each of these programs were considered for further analysis. In the docking of ssDNA coated RecA with AlkB, the output complex from PatchDock has a Score and ACE value of 16260 and 35.25 respectively, which was further subjected to FireDock for post energy minimization. The Global Energy value and Attractive Vander Waals energy after Firedock analysis were -21.47Kcal/mol and -38.38Kcal/mol respectively. Analysis with ClusPro docking tool gave the clustering scores of -972.0 with RecA and AlkB. It was observed that the bases in the RecA coated ssDNA are in close proximity with nucleotide recognition lid of the AlkB protein. The AlkB protein adopts an energetically favorable conformation and interacts without any steric hindrance. Since the docking methodology employed involves the rigid protein-protein docking, the flexibility of the loop regions of the AlkB molecules were not considered. The non-covalent interactions of the RecA and SSB coated ssDNA with AlkB was monitored using Protein-Protein interaction analysis tool from Schrodinger. It was noted that the residues Gln84, Leu130, Thr56, Tyr55 and Arg161 of AlkB interacts with Lys245, Met202, Gly20, Arg169 and Arg161 residues of RecA protein. In ZDOCK it was observed that Phe196, Tyr76 and Lys 134 shows electrostatic interaction with Thy5, Thy4 and Thy3 respectively. The interatomic distances between the bases and the AlkB protein was studied using the WhatIf server. Thus, we assume that AlkB is recruited through the protein-protein interaction with SSB and RecA which provides a stable platform for AlkB to interact with ssDNA thereby enhancing its DNA repair activity.

Docking of AlkB with RecA filament bound ssDNA using ClusPro

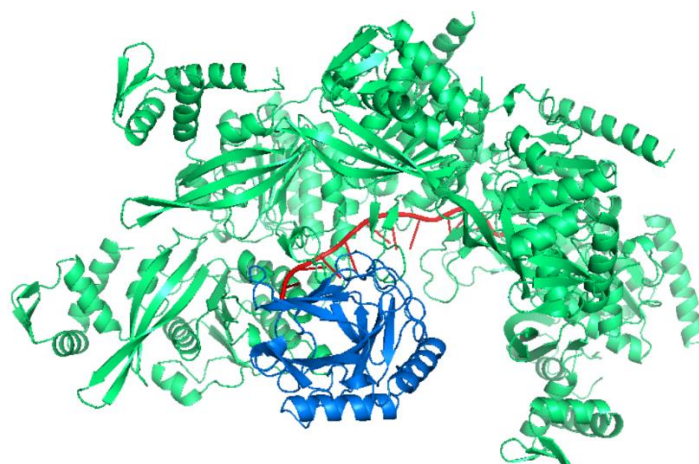
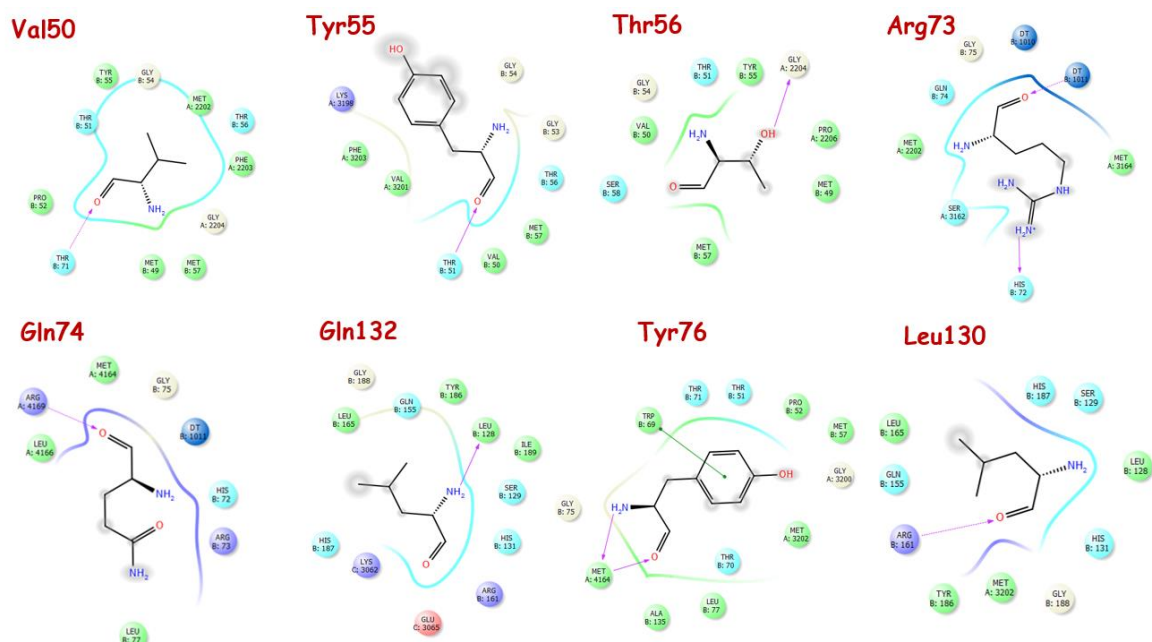


Figure 15: Docked pose of AlkB with RecA filament bound ssDNA generated using ClusPro. SSB tetramer, AlkB and ssDNA are represented in green, blue and orange colors respectively.

Figure 16: 2D interaction diagram of AlkB with RecA filament



Tabular column 4: Residues of AlkB interacting with RecA filament

S.No	AlkB residues	SSB residues	Chain	Distance Å	Nature of interaction
1	Val50	Met 202	A	3.61	Hydrophobic
2	Arg73	Met202	A	3.27	Hydrophobic
		Thy10	A	3.21	Electrostatic
		Thy11	A	3.15	Electrostatic
3	Thr56	Gly204	A	3.33	Hydrogen bonding
4	Tyr55	Val201	A	3.36	Hydrophobic
5	Leu130	Met202	A	3.77	Hydrophobic
6	Tyr76	Lys198	A	3.28	Electrostatic
		Met164	A	3.24	Hydrogen bonding
		Pro206	A	3.05	Hydrophobic
7	Gln74	Arg169	A	3.60	Hydrogen bonding
		Thy11	A	3.42	Electrostatic

Docking of AlkB with RecA filament bound ssDNA using PatchDock

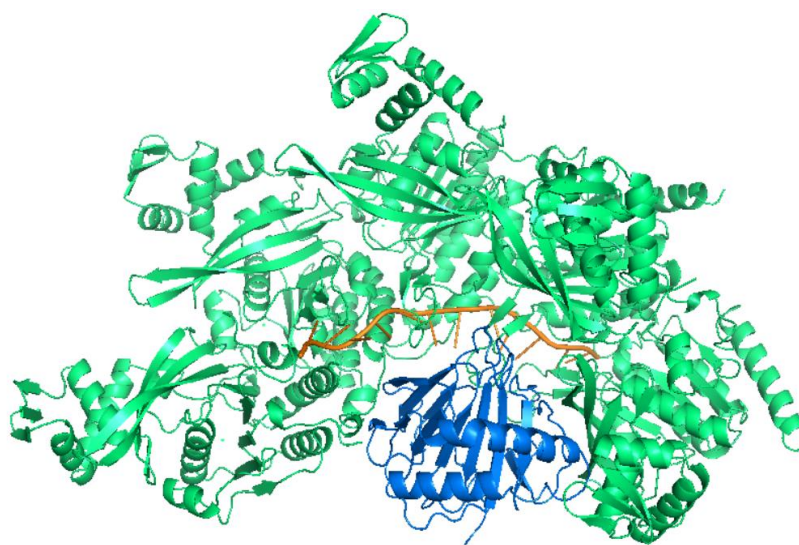
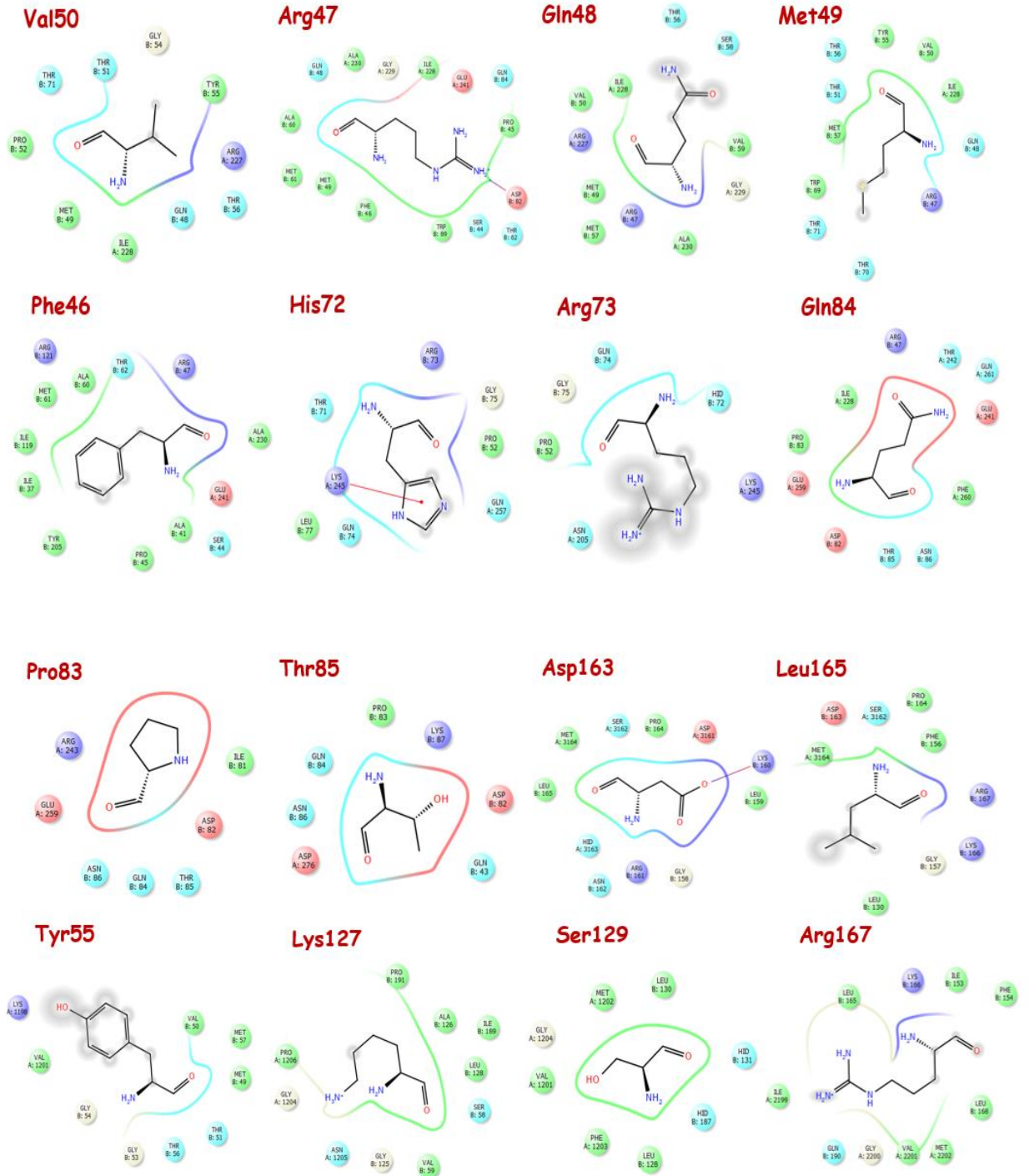


Figure 17: Docked pose of AlkB with RecA filament bound ssDNA generated using ClusPro. SSB tetramer, AlkB and ssDNA are represented in green, blue and orange colors respectively.

Figure 18: 2D interaction diagram of AlkB with RecA filament



Tabular column 5: Residues of AlkB interacting with RecA filament

S.No	AlkB residues	SSB residues	Chain	Distance Å	Nature of interaction
1	Val50	Arg227	A	3.62	Hydrophobic
2	Arg47	Ile228	A	3.11	Electrostatic Hydrophobic
3	Met49	Ile228	A	3.50	Hydrophobic Electrostatic
4	His72	Lys215	A	3.21	Cation- π interaction
5	Phe46	Ala230	A	3.38	Electrostatic
6	Arg121	Lys232	A	3.60	Electrostatic
7	Gln84	Lys245 Thr242, Gln261	A A,A	3.07 2.61	Electrostatic Polar
8	Pro83	Glu259	A	3.64	Electrostatic
9	Tyr55	Lys198	A	3.44	Electrostatic
10	Leu165	Ser162	A	2.47	Hydrophobic

Docking of AlkB with RecA filament bound ssDNA using ZDock

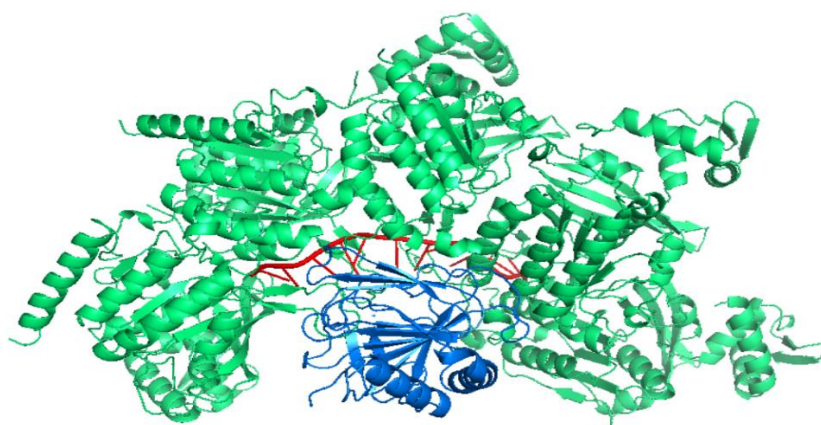
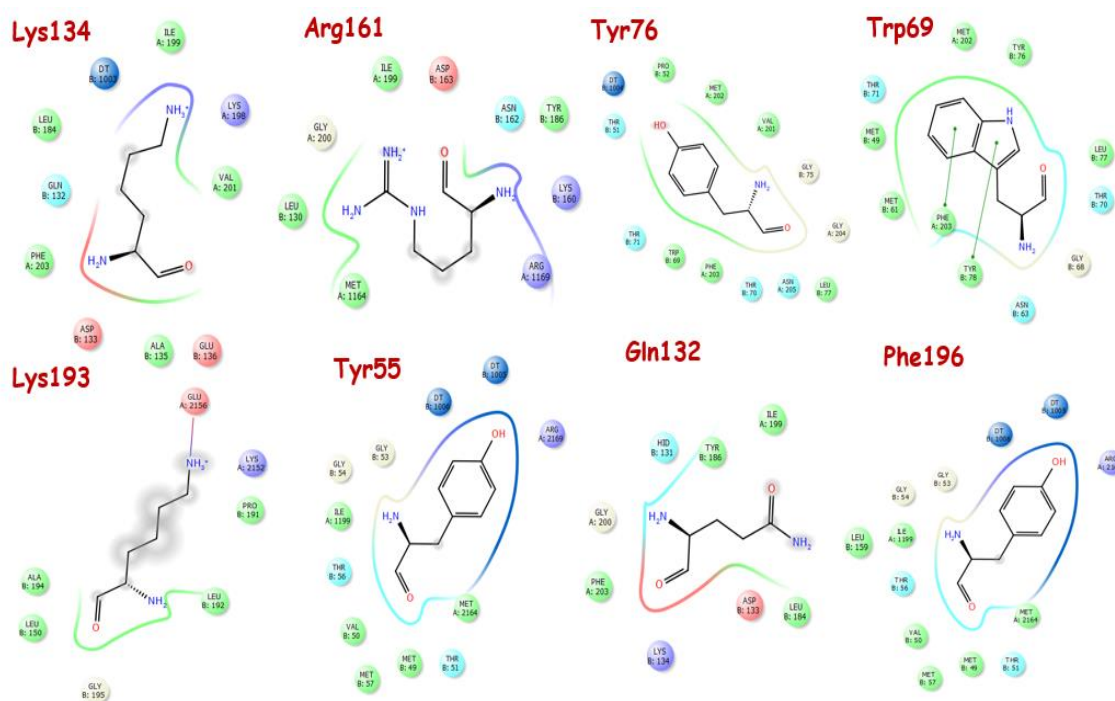


Figure 19: Docked pose of AlkB with RecA filament bound ssDNA generated using ClusPro. SSB tetramer, AlkB and ssDNA are represented in green, blue and orange colors respectively.

Figure 20: 2D interaction diagram of AlkB with RecA filament



Tabular column 6: Residues of AlkB interacting with RecA filament

S.No	AlkB residues	SSB residues	Chain	Distance Å	Nature of interaction
1	Lys134	Lys198 Val201	A A	3.17 3.33	Electrostatic Hydrophobic
2	Arg161	Ile199 Met164	A A	3.26 3.65	Hydrophobic Hydrophobic
3	Lys193	Glu156	A	2.65	Salt bridge
4	Tyr76	Val201 Met202	A A	3.71 3.57	Hydrophobic Electrostatic
5	Trp69	Phe203	A	5.18	π - π interaction
6	Leu77	Phe203	A	3.14	Hydrophobic
7	Gly53	Gly200	A	3.28	Hydrophobic
8	Lys127	Leu166	A	3.50	Hydrophobic
9	Tyr55	Arg169 Thy6, Thy5	A B,B	3.36 3.62	Electrostatic Electrostatic
10	Phe196	Met164 Thy5, Thy6	A B,B	2.97 3.18	Electrostatic

Docking of AlkB with RecA filament

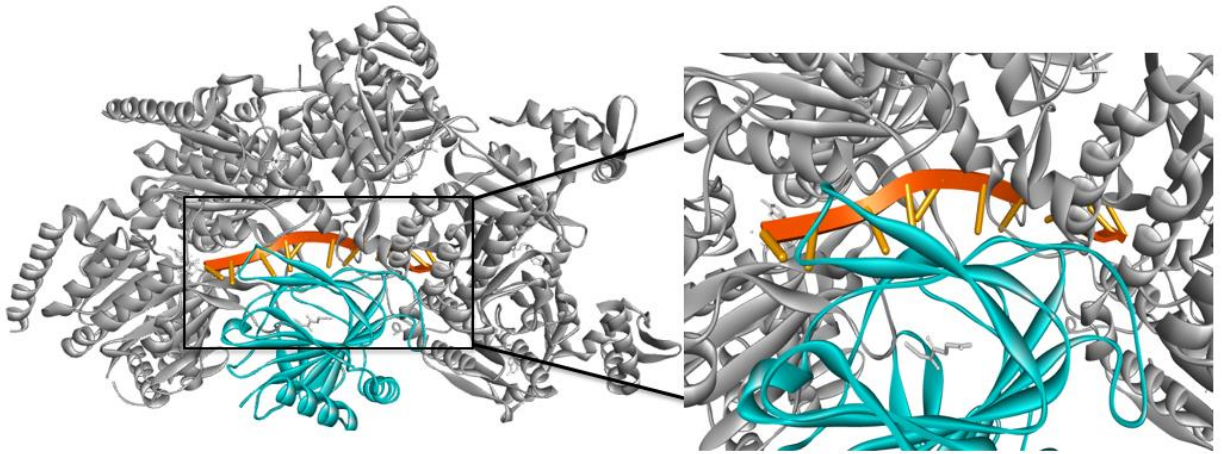


Figure 21: Model illustrating the interaction of AlkB with ssDNA bound RecA filament. *In silico* docking analysis using ZDOCK shows that AlkB could dock in close proximity with ssDNA without any steric hindrance and adopt an energetically favorable conformation. The AlkB protein and RecA filament are represented in cyan and grey color cartoon respectively. The sugar phosphate backbone of ssDNA are represented in orange color.

References

- 1) Rajski, S. R., and Williams, R. M. (1998) DNA Cross-Linking Agents as Antitumor Drugs. *Chem.Rev.* 98, 2723–2796.
- 2) Beranek, D. T. (1990) Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutat. Res.* **231**, 11–30
- 3) Rydberg, B. and T. Lindahl, Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl- -methionine is a potentially mutagenic reaction. *EMBO J.* , 1982. 1 p. 211–216.
- 4) Rajski, S. R., and Williams, R. M. (1998) DNA Cross-Linking Agents as Antitumor Drugs. *Chem.Rev.* 98, 2723–2796.
- 5) Beranek, D. T. (1990) Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutat. Res.* **231**, 11–30
- 6) Johnson, R. E., Yu, S. L., Prakash, S., and Prakash, L. (2007) A role for yeast and human translesion synthesis DNA polymerases in promoting replication through 3-methyl adenine. *Mol. Cell. Biol.* 27, 7198–7205
- 7) Kunz, B. A., Straffon, A. F., and Vonarx, E. J. (2000) DNA damage-induced mutation: tolerance via translesion synthesis. *Mutat. Res.* 451, 169–185
- 8) Washington, M. T., Johnson, R. E., Prakash, S., and Prakash, L. (2000) Accuracy of thymine-thymine dimer bypass by *Saccharomyces cerevisiae* DNA polymerase η . *Proc. Natl. Acad. Sci. U.S. A.* 97, 3094–3099
- 9) Acharya, N., Johnson, R. E., Prakash, S., and Prakash, L. (2006) Complex formation with Rev1 enhances the proficiency of *Saccharomyces cerevisiae* DNA polymerase ζ

for mismatch extension and for extension opposite from DNA lesions. *Mol. Cell. Biol.* 26, 9555–9563

10) Falnes, P. O., Bjoras, M., Aas, P. A., Sundheim, O., and Seeberg, E. (2004) Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res.* 32, 3456–3461

11) Falnes, P. O. (2004) Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Res.* 32, 6260–6267

12) Delaney, J. C., and Essigmann, J. M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in *alkB* *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14051–14056

13) Falnes, P. O., Klungland, A., and Alseth, I. (2007) Repair of methyl lesions in DNA and RNA by oxidative demethylation. *Neuroscience* 145, 1222–1232

14) Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., and Sedgwick, B. (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* 419, 174–178

15) Aas, P. A., Otterlei, M., Falnes, P. O., Vagbo, C. B., Skorpen, F., Akbari, M., Sundheim, O., Bjoras, M., Slupphaug, G., Seeberg, E., and Krokan, H. E. (2003) Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* 421, 859–863

16) Ougland, R., Zhang, C. M., Liiv, A., Johansen, R. F., Seeberg, E., Hou, Y. M., Remme, J., and Falnes, P. O. (2004) AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. *Mol. Cell* 16, 107–116

- 17) Kurowski, M. A., Bhagwat, A. S., Papaj, G., and Bujnicki, J. M. (2003) Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics* 4, 48
- 18) Gerken, T., Girard, C. A., Tung, Y. C., Webby, C. J., Saudek, V., Hewitson, K. S., Yeo, G. S., McDonough, M. A., Cunliffe, S., McNeill, L. A., Galvanovskis, J., Rorsman, P., Robins, P., Prieur, X., Coll, A. P., Ma, M., Jovanovic, Z., Farooqi, I. S., Sedgwick, B., Barroso, I., Lindahl, T., Ponting, C. P., Ashcroft, F. M., O’Rahilly, S., and Schofield, C. J. (2007) The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* (80-.). 318, 1469–1472
- 19) Sedgwick, B., Bates, P. A., Paik, J., Jacobs, S. C., and Lindahl, T. (2007) Repair of alkylated DNA: recent advances. *DNA Repair (Amst)*. 6, 429–442
- 20) Wei, Y. F., Chen, B. J., and Samson, L. (1995) Suppression of *Escherichia coli* alkB mutants by *Saccharomyces cerevisiae* genes. *J. Bacteriol.* 177, 5009–5015
- 21) Choudhary, V., and Schneiter, R. Pathogen-Related Yeast (PRY) proteins and members of the CAP superfamily are secreted sterol-binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16882–16887
- 22) Verna, J., Lodder, A., Lee, K., Vagts, A., and Ballester, R. (1997) A family of genes required for maintenance of cell wall integrity and for the stress response in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 13804–13809
- 23) Aravind, L., and Koonin, E. V (2001) The DNA-repair protein AlkB, EGL-9, and Iprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* 2, RESEARCH0007

- 24) Keeling, K. M., Salas-Marco, J., Osherovich, L. Z., and Bedwell, D. M. (2006) Tpa1p is part of an mRNP complex that influences translation termination, mRNA deadenylation, and mRNA turnover in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 26, 5237–5248
- 25) Henri, J., Rispal, D., Bayart, E., van Tilbeurgh, H., Seraphin, B., and Graille, M. (2010) Structural and functional insights into *Saccharomyces cerevisiae* Tpa1, a putative prolylhydroxylase influencing translation termination and transcription. *J. Biol. Chem.* 285, 30767–30778
- 26) Kim, H. S., Kim, H. L., Kim, K. H., Kim do, J., Lee, S. J., Yoon, J. Y., Yoon, H. J., Lee, H. Y., Park, S. B., Kim, S. J., Lee, J. Y., and Suh, S. W. (2009) Crystal structure of Tpa1 from *Saccharomyces cerevisiae*, a component of the messenger ribonucleoprotein complex. *Nucleic Acids Res.* 38, 2099–2110
- 27) Loenarz, C., Sekirnik, R., Thalhammer, A., Ge, W., Spivakovsky, E., Mackeen, M. M., McDonough, M. A., Cockman, M. E., Kessler, B. M., Ratcliffe, P. J., Wolf, A., and Schofield, C. J. (2014) Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4019–4024
- 28) Hanway, D., Chin, J. K., Xia, G., Oshiro, G., Winzeler, E. A., and Romesberg, F. E. (2002) previously uncharacterized genes in the UV- and MMS-induced DNA damage response in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10605–10610

29) Discovery Studio Modeling Environment, Release 3.5. Accelrys Software Inc.: San Diego; 2012.

30) Sastry, G. Madhavi, et al. "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments." *Journal of computer-aided molecular design* 27.3 (2013): 221-234.

31) Søndergaard, Chresten R., et al. "Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values." *Journal of Chemical Theory and Computation* 7.7 (2011): 2284-2295.

32) Sastry, G. Madhavi, et al. "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments." *Journal of computer-aided molecular design* 27.3 (2013): 221-234.

33) Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Computational Chemistry* 2009, 16: 2785-91.

34) Friesner, Richard A., et al. "Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes." *Journal of medicinal chemistry* 49.21 (2006): 6177-6196.

35) Cho, A. E.; Guallar, V.; Berne, B.; Friesner, R. A., "Importance of Accurate Charges in Molecular Docking: Quantum Mechanical/Molecular Mechanical (QM/MM) Approach," *J. Comput. Chem.*, 2005, 26, 915–931

- 36) Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7, e46688
- 37) Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.* 33, W382–8
- 38) ö Bochevarov, A.D.; Harder, E.; Hughes, T.F.; Greenwood, J.R.; Braden, D.A.; Philipp, D.M.; Rinaldo, D.; Halls, M.D.; Zhang, J.; Friesner, R.A., "Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences," *Int. J. Quantum Chem.*, 2013, 113(18), 2110-2142
- 39] Hecht, Stephen S., and Dietrich Hoffmann. "Tobacco-specific nitrosamines, an important group of carcinogens in tobacco and tobacco smoke." *Carcinogenesis* 9.6 (1988): 875-884.
- 40] Trewick, Sarah C., et al. "Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage." *Nature* 419.6903 (2002): 174-178.
- 41] Lindahl, T., B. Sedgwick, M. Sekiguchi, and Y. Nakabeppu. 1988. Regulation and expression of the adaptive response to alkylating agents. *Annu. Rev. Biochem.* 57:133–157.
- 42] Dosanjh, Manjit K., B. Singer, and John M. Essigmann. "Comparative mutagenesis of O6-methylguanine and O4-methylthymine in *Escherichia coli*." *Biochemistry* 30.28 (1991): 7027-7033.

- 43] Slesarev, Alexei I., et al. "The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens." *Proceedings of the National Academy of Sciences* 99.7 (2002): 4644-4649.
- 44] Falnes, P. Ø., A. Klungland, and I. Alseth. "Repair of methyl lesions in DNA and RNA by oxidative demethylation." *Neuroscience* 145.4 (2007): 1222-1232.
- 45] Yang, Cai-Guang, et al. "Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA." *Nature* 452.7190 (2008): 961-965.
- 46] Yu, Bomina, et al. "Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB." *Nature* 439.7078 (2006): 879-884.
- 47] Holland, Paul J., and Thomas Hollis. "Structural and mutational analysis of *Escherichia coli* AlkB provides insight into substrate specificity and DNA damage searching." *PLoS One* 5.1 (2010): e8680.
- 48] Meyer, Ralph R., and PHYLLIS S. Laine. "The single-stranded DNA-binding protein of *Escherichia coli*." *Microbiological reviews* 54.4 (1990): 342.
- 49] Raghunathan, Srinivasan, et al. "Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-Å resolution." *Proceedings of the National Academy of Sciences* 94.13 (1997): 6652-6657.
- 50] Raghunathan, Srinivasan, et al. "Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA." *Nature structural biology* 7.8 (2000): 648-652.

- 51] Story, Randall M., and Thomas A. Steitz. "Structure of the recA protein-ADP complex." (1992): 374-376.
- 52] Chen, Zhucheng, Haijuan Yang, and Nikola P. Pavletich. "Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures." *Nature* 453.7194 (2008): 489-494.
- 53] Weiner, JOEL H., LEROY L. Bertsch, and A. Kornberg. "The deoxyribonucleic acid unwinding protein of Escherichia coli. Properties and functions in replication." *Journal of Biological Chemistry* 250.6 (1975): 1972-1980.
- 54] Duhovny D, Nussinov R, Wolfson HJ. Efficient Unbound Docking of Rigid Molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag, 2002.
- 55] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucl. Acids. Res.* 33: W363-367, 2005.
- 56] Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. (2014) ZDOCK Server: Interactive Docking Prediction of Protein-Protein Complexes and Symmetric Multimers. *Bioinformatics* 30(12): 1771-3.
- 57] Kozakov D, Beglov D, Bohnuud T, Mottarella S, Xia B, Hall DR, Vajda, S. How good is automated protein docking? *Proteins: Structure, Function, and Bioinformatics*, 2013 Aug

