# PRODUCTION AND PURIFICATION OF MUTL GENE AND UVRD RECOMBINANT PROTEIN FOR HELICASE DEPENDENT DNA AMPLIFICATION

**KALAIVANI BUSKARAN**

**UNIVERSITI SAINS MALAYSIA**
**2015**

# PRODUCTION AND PURIFICATION OF MUTL GENE AND UVRD RECOMBINANT PROTEIN FOR HELICASE DEPENDENT DNA AMPLIFICATION

**by**

**KALAIVANI BUSKARAN**

**Thesis submitted in fulfilment of the requirements for the degree of Master of Science**

**August 2015**

# ACKNOWLEDGEMENTS

First and foremost all praise and glory to Almighty God for His grace and help in all my endeavours. My family, without whose encouragement, I would not have been here. I thank them for their unflinching prayers and support.

My profound gratitude and appreciation goes to my supervisor, Associate Professor Dr. Sasidharan Sreenivasan. His invaluable help of constructive comments and suggestions throughout the experimental and thesis works have contributed to the success of this research. Not forgotten, my appreciation to my co-supervisor, Dr. Mervyn Liew Wing On and Dr. Eugene Ong Beng for their support and valuable knowledge regarding this topic.

I would like to express my appreciation to the Institute for Research in Molecular Medicine (INFORMM) for their support and help towards my postgraduate affairs. My acknowledgement also goes to all the staffs of School of Biological Sciences, USM for their co-operations and supports. Sincere thanks to all my friends especially and others for their kindness and moral support during my study. Thanks for the friendship and memories.

Lastly, I would like to express my gratitude to The Malaysian Ministry of Higher Education, MyMaster scholarship and Graduate Assistant scholarship from USM for their financial support and also USM Research University Cluster (RUC) Grant (1001/PSKBP/86300110), USM for funding of this research.

## TABLE OF CONTENTS

**Page**

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| °C | degree Celsius |
| µl | micro litre |
| APS | Ammonium persulfate |
| ATP | Adenosine triphosphate |
| ATPase | Adenosinetriphophatase |
| BLAST | Basic Local Alignment Search Tool |
| Bp | Base pair |
| BSA | Bovine serum albumin |
| $CaCl_2$ | Calcium Chloride |
| $ddH_2O$ | double-distilled water |
| DNA | deoxyribonucleic acid |
| dNTP | deoxyribonucleotide triphosphate |
| dsDNA | double-stranded deoxyribonucleic acid |
| DTT | dithiothreitol |
| *E. coli* | *Escherichia coli* |
| EDTA | Ethylenediaminetetraacetic acid |
| g | gram |
| g | gravity |
| h | hour |
| HCl | hydrochloric acid |
| HDA | helicase-dependent amplification |
| IMAC | Immobilized metal ion affinity chromatography |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| $K_2HPO_4$ | Dipotassium hydrogen phosphate |
| Kb | kilo bases |
| Kbp | kilo base pairs |
| KCl | Potassium chloride |
| kDa | kilo Dalton |
| $KH_2PO_4$ | Potassium dihydrogen phosphate |
| kV | kilo Voltage |
| L | Liter |
| LB | Luria-Bertani |
| MALDI-TOF/TOF | Matrix Assisted Laser Desorption Ionization-Time of Flight/Time of Flight |
| $MgCl_2$ | Magnesium chloride |
| Min | Minute |
| mL | milli litre |
| mM | milli molar |

| | |
|---|---|
| MWCO | Molecular weight cut off |
| NaCl | Sodium Chloride |
| NCBI | National Center for Biotechnology Information |
| $NH_{42}SO_4$ | Ammonium sulfate |
| Ni-NTA | Nickel-nitrilotriacetic acid |
| nm | Nanometer |
| NMR | Nuclear magnetic resonance |
| OD | Optical Density |
| PBS | Phosphate Buffered Saline |
| PCR | Polymerase Chain Reaction |
| POC | Point-of-care |
| RNA | Ribonucleic acid |
| rpm | Revolutions per minute |
| SB | Super broth |
| SDS | Sodium dodecyl sulfate |
| SDS-PAGE | Sodiumdodecylsulfate-polyacrylamide gel electrophoresis |
| sec(s) | second(s) |
| SSB | single-strand binding |
| ssDNA | single-stranded deoxyribonucleic acid |
| TB | Terrific broth |
| TBE | Tris-Borate-EDTA |
| TBS | Tris-buffered saline |
| TBS-T | Tris-buffered saline-Tween 20 |
| TE | Tris-EDTA |
| TEMED | Tetramethylethylenediamine |
| Tm | Melting temperature |
| U/l | Units per litre |
| UV | Ultraviolet |
| V | Volt |

# PENGHASILAN DAN PENULENAN MUTL GEN DAN UVRD REKOMBINAN PROTEIN UNTUK UJIAN AMPLIFIKASI DNA YANG BERSANDARKAN HELICASE

## ABSTRAK

Kajian ini telah dijalankan untuk mengoptimumkan penghasilan rekombinan UvrD helicase dan rekombinan MutL protein untuk ujian amplifikasi helicase. Rekombinan UvrD helicase telah diklon dan diekspresikan dengan vektor Lemo21. Penghasilan UvrD helicase protein telah dioptimumkan 10% dengan menggunakan kaedah konvensional kelalang bergoncang. Penghasilan rekombinan UvrD helicase yang optimum telah berjaya dihasilkan dengan kaldu Terrific, 0.2 mM kepekatan IPTG, 9 jam masa induksi dan pada suhu 37°C selepas induksi. Penghasilan rekombinan UvrD/pET28a/Lemo21 telah dioptimumkan dengan teknik fermentasi menggunakan kelalang bergoncang skala kecil memberikan hasil protein UvrD rekombinan sebanyak 1.35 ng/µg dibandingkan dengan penghasilan 0.143 ng/µg dari kajiann awal yang dijalankan. Disamping, rekombinan UvrD helicase berjaya befungsi rekombinan MutL protein telah diasingkan dan diamplifikasikan daripada pencilan *Escherichia coli* tempatan. Tindak balas Berantai Polimerase mangamplifikasikan MutL protein dengan saiz 1848 bp. MutL gen kemudian diklon ke dalam vektor pengklonan TOPO PCR 2.1. Klon yang berjaya kemudiannya telah dikaji menggunakan urutan Uniprot dan didapati mengandungi 615 asid amino yang menunjukkan 100% persamaan dengan MutL protein daripada spesies *E. coli* lain. Kemudian, MutL gen telah diklon ke dalam vektor pET/ 28a (+) dan diekspresikan dengan menggunakan vektor Lemo21. Rekombinan UvrD dan MutL telah ditulenkan

dengan menggunakan kaedah kromatografi afiniti logam tanpa menyebabkan penyahaslian protein. Keputusan kajian SDS-PAGE mengesahkan kehadiran protein yang dihasilkan. Immunoreaktiviti rekombinan didapati sensitif dan khusus apabila diuji dengan kaedah analisis pemblotan Western. Kajian MALDI TOF / TOF juga dijalankan untuk mengesahkan kehadiran protein pada saiz 82 kDa dan 68 kDa bagi protein rekombinan UvrD helicase dan MutL gen masing-masing. Kemudian, aktiviti rekombinan UvrD helicase telah disahkan dengan menggunakan ujian aktiviti penguraian oleh helicase secara *in vitro* dan ujian amplifikasi yang bersandarkan helicase. Kajian aktiviti penguraian oleh helicase secara *in vitro* berdasarkan masa khusus dengan menggunakan kepekatan yang berbeza (250 ng dan 500 ng) menunjukkan reaksi aktiviti enzim positif. Akhir sekali, ujian HDA telah mengesahkan keupayaan UvrD helicase rekombinan dan MutL protein dapat mengurai jujukan tertentu pada DNA dupleks. Kesimpulannya, objektif kajian ini telah dicapai dengan mengoptimumkan dan menghasilkan UvrD helicase rekombinan dan MutL gen rekombinan yang telah disahkan aktiviti protein tersebut melalui ujian aktiviti penguraian oleh helicase secara *in vitro* dan ujian amplifikasi yang bersandarkan helicase.

# PRODUCTION AND PURIFICATION OF MUTL GENE AND UVRD RECOMBINANT PROTEIN FOR HELICASE DEPENDENT DNA AMPLIFICATION

## ABSTRACT

This research was conducted to optimize yield of recombinant UvrD helicase and to produce recombinant MutL protein for the helicase amplification dependent reaction. Recombinant UvrD helicase were cloned and expressed into Lemo21 expression host. The overall yield of the UvrD helicase protein was improved 10% using conventional shake flask cultivation method. The optimum recombinant UvrD helicase were successfully produced in terrific broth culture medium, at 0.2 mM IPTG inducer concentration, 9 hours of post-induction time and 37°C post-induction temperature. The cultivation of the recombinant UvrD/pET28a/Lemo21 in small scale fermentation using shake flask culture has optimized the yield to 1.35 ng/µg recombinant UvrD helicase proteins as compared to 0.143 ng/µg of recombinant UvrD helicase protein from previous studies. In order for, recombinant UvrD helicase function, recombinant MutL protein was produce from local isolate of *Escherichia coli*. The PCR amplified MutL protein had size of 1848 bp of sequence analysis. The MutL gene was then cloned into TOPO PCR 2.1 cloning vector. The successful clone was then analysed using Uniprot software produce 615 deduced amino acid 100% similarities to MutL protein from other interspecific species of *E. coli*. Later, the amplified MutL gene was cloned into pET/28a(+) expression vector and expressed in Lemo21 expression host. The recombinant UvrD and MutL was purified under non-denaturing conditions using immobilized metal affinity chromatography. SDS-PAGE was ran and confirmed the presence of the protein

expression. Immunoreactivity of the recovered recombinant proteins was found to be sensitive and specific when tested with Western blot. MALDI TOF/TOF also conducted to confirm the recombinant protein at expected at 82 kDa and 68 kDa for UvrD helicase and MutL protein respectively. Subsequently, the activity of recombinant UvrD helicase was validated using unwinding helicase activity and helicase dependent amplification assay. The time course analysis of *in vitro* unwinding helicase activity at different concentration (250 ng and 500 ng) had shown a positive enzyme unwinding activity. Finally, the HDA reaction confirmed the ability of the in-house recombinant UvrD helicase and recombinant MutL gene to unwind a specific sequence from duplex DNA. In summary, the objectives of this research has been achieved successfully by optimizing and producing recombinant UvrD helicase and recombinant MutL protein which were validated by using *in vitro* unwinding helicase activity and helicase dependent amplification assay.

# CHAPTER 1.0: INTRODUCTION

## 1.1 Overview

Helicases are motor protein that bind and remodel nucleic acid or nucleic acid protein complexes. The first DNA helicase was found *E. coli* in 1976. The prominent role of helicase is to separate the double-stranded (ds) DNA into single strands and allow each strand to be copied (Anonymous, 2014). Helicase disrupts this hydrogen bond by hydrolysis reaction of a nucleoside 5' -triphosphate (NTP) which serves as the energy currency of cells. Thus, helicases are DNA-dependent nucleoside 5' triphosphatases (NTPases) (Matson *et al*., 1990). Helicases function in processes including DNA replication, repair, recombination and bacterial conjugation, and are a component of eukaryotic transcription complexes (Timothy, 2004).

Helicase with polymerases and other accessory proteins has developed a Helicase dependent amplification (HDA) which a new molecular diagnostic technology of Nucleic acid amplification. HDA utilizes the activity of a DNA helicase to separate complementary strands from double-stranded nucleic acids. DNAs evade the temperature cycling to form single stranded DNA templates for primer hybridization and followed by primer extension by a DNA polymerase (Vincent *et al*., 2004). The helicase enzyme plays a role as substitutes for the high temperature denaturation process held in PCR and allows isothermal DNA amplification. This isothermal test has been used to amplify the target sequences at single temperature from bacteria without initial heat denaturation. Using helicase to denature the target DNA allows the HDA platform to be a truly isothermal amplification technic (An *et al*., 2005).

This study is a continuation of an in-house UvrD helicase (82 kDa) from *Escherichia coli* that was produced by a colleague (Yamuna *et al*., 2012). The UvrD helicase capable of melting fully duplex molecules, blunt-ended DNA fragments as well as nicked circular DNA. UvrD helicase plays a major role in methyl-directed mismatch repair (Modrich, 1991), DNA excision repair (Sancar, 1996), replication restart and plasmid replication.

The helicase dependent amplification tests were developed using UvrD helicase and accessory proteins (Single Strand Binding (SSB) protein, MutL and T4 gene). SSBs specifically attach to the single stranded of DNA in a sequence-independent manner. They intensify polymerase activity by preserving unwound ssDNAs from degradation and by destabilizing DNA secondary structure (Laura *et al*., 2013). UvrD helicase mediated amplification requires MutL. MutL is the master coordinator of mismatch repair in living organism. It enrols UvrD helicase to unwind the DNA strand containing the replication error. MutL stimulates UvrD helicase activity more than tenfold by loading it onto the DNA substrate (Vincent *et al*., 2004). In addition to recombinant UvrD helicase production optimization, recombinant MutL protein, which is not available commercially but necessary for the HDA test was also produced in this study.

The UvrD helicase production was conducted using the following parameters: the different bacterial culture medium, inducer concentration trial, induction time trial and post-induction temperature for optimum production of recombinant UvrD helicase using shake flask system. On the other hand, MutL protien was identified, cloned, expressed and purified from a local isolate. The both recombinant UvrD helicase and recombinant MutL protien were used for HDA activity.

## 1.2 Statement of Problem and Rationale of Study

A major challenge for the medical community is to develop diagnostic tests that meet the needs of the people from poor resource health care facilities; majorities are from the developing and third world countries. Several factors such as antibiotic resistance, cost of effective drugs, and increased threat of an accelerated epidemic-to-pandemic transition of a communicable disease have increased the support and attention from the health sector on the need for improved diagnostic technologies (Paul *et al*., 2008).

In this rapid growth of diagnostic field, the developments of HDA assay becoming prominent. Unlike polymerase chain reaction PCR, HDA uses a helicase enzyme, rather than heat, to separate double-stranded DNA. This allows DNA amplification without the need for thermo cycler, which is used for *in vitro* DNA amplification purposes. Like PCR, the HDA reaction selectively amplifies a target sequence defined by two primers; enabling primer annealing and extension by a strand–displacing DNA polymerase (Vincent *et al*., 2004).

There are certain shortcomings of the PCR-based methods. It requires thermo cycling to separate two DNA strands (An *et al*., 2005). PCR assay requires expensive instrumentation and well-trained personnel for operation. This platform uses very large and expensive instrumentation. The reagent and consumable costs are from USD$ 12 to 14 (RM 42 to 49) per specimen, and approximately 1 to 2 hands-on techician and 6 to 8 hours for the total turnaround time (Paul *et al*., 2008).

The commercially available isothermal DNA helicase costs USD$ 568.00 (about RM 2000) for 20 µg from Biohelix Corporation has become a major problem for the diagnostic test especially in poor countries. Therefore, production of an in-house

DNA helicase is a practical solution for various applications including the development of a non-invasive and inexpensive PCR-based diagnostic kit.

Development of a simple, cheap, rapid diagnostic, sensitive and easy-to-adapt test for accurate identification of pathogenic strains in a timely fashion is highly desirable for hospital cost containment, patient management, and prompts epidemiological interventions (Kim and Easley, 2011). Helicase dependent amplification technology is developed as a hand held diagnostic device that facilitates pathogen detection in the field at point of care. Also, time to diagnose disease is reduced. Earlier diagnosis leads to earlier initiation of treatment, a reduced period of infectiousness, and improved patient outcomes (Yamuna *et al.*, 2012).

## 1.3 Objectives

The current study was undertaken with the following objectives:

1. To clone and express recombinant UvrD helicase gene in pET28a(+) with Lemo21 expression vector.
2. To clone and express recombinant MutL gene in pET28a(+) with Lemo21 expression vector.
3. To optimize the protein expression conditions in a shake-flask system and purification of recombinant helicase.
4. To verify the recombinant helicase activity by using *in vitro* helicase activity assay and Helicase Dependent Amplification (HDA) assay.

# CHAPTER 2.0: LITRATURE REVIEW

## 2.1 Helicase Overview

Helicase was first discovered in *Escherichia coli* in 1976, since then RNA and DNA helicases with diverse functions have been found in all organisms. DNA helicases are enzymes capable of unwinding double-stranded DNA to provide the single-stranded template required in many biological events, such as replication, recombination and repair (Soultanas and Wigley, 2001). The enzyme commission number of this DNA helicase is 3.6.4.12 and it is grouped in hydrolase since it utilizes the energy from ATP hydrolysis to unwind double-stranded DNA (George *et al.*, 2009).

### 2.1.1 Function of helicases

Helicase is grouped into hydrolase category where they couple the chemical energy ATP binding and hydrolysis to the unwinding of duplex DNA or RNA into single-strands. After helicase split the hydrogen bonds between the duplex helix of DNA and dislocate other non-covalent connections between complementary base pairs, it will move unidirectionally next to the bound strand. Helicases exhibit these activities in isolation, but work competently as part of a larger protein complex. So, helicases perform as a model system to understand ATPase-coupled motors. All helicases are translocases and DNA-dependent ATPases. DNA helicases contain at least three common biochemical properties which are DNA binding, ATP (or other nucleotides) hydrolysis, and ATP-dependent DNA unwinding activity.

## 2.2 UvrD Helicase

UvrD helicase is the most abundant type of helicase in *E. coli.* The UvrD protein consists of 720 amino acids with molecular mass of approximately 82 kDa. This helicase is from member of the SF1 helicase superfamily. The UvrD helicase is capable of melting fully duplex molecules (DNA fragment with blunt ends) as well as nicked circular DNA molecules (Matson *et al*., 1990). Its major functions are in methyl-directed mismatch repair (Modrich, 1991), DNA excision repair (Sancar, 1996), replication restart (Flores *et al*., 2004, 2005; Michel *et al*., 2004), and plasmid replication (Bruand and Ehrlich, 2000), and it can also take apart RecA protein filaments formed on ssDNA (Veaute *et al*., 2005), most probably by displacing RecA from ssDNA telomerase from telomeric DNA ends (Boule *et al*., 2005).

UvrD is an important enzyme in the DNA repair and it's involved in both mismatch repair pathway and nucleotide excision repair. With its unfailing multiple roles, the UvrD deficient cells are more susceptible to DNA-damaging agents and unveils an increase level of the mutation and are hyper-recombinogenic (Washburn *et al*., 1991). Disruption of MMR leads to increased spontaneous mutations and homologous recombination, whereas disruption of NER increases sensitivity to UV and other agents of DNA damage.

In the mismatch repair pathway, UvrD initiates by unwinding at the d(GATC)-located nick created by MutH and, together with an appropriate exonuclease, facilitates removal of the unmethylated daughter strand containing the mismatch. Direct physical interactions between UvrD and MutL, the master coordinator of the mismatch repair pathway, have shown that MutL dramatically stimulates UvrD helicase activity incision by the combined action of MutS, MutL

and MutH. Disruption on MMR leads to increased spontaneous mutations and homologous recombination (Yamaguchi *et al*., 1995).

UvrD also participates in the UvrABC nucleotide excision repair pathway by removing the 12–13 base oligonucleotide containing a pyrimidine dimer or bulky adduct. The NER pathway needs UvrD to unwind the removed oligonucleotide that compose the damaged nuclei acid base pairs and expel UvrC, with resynthesis by DNA polymerase I (Caron *et al*., 1985). The mismatch repair activity uses UvrD to assists MutL at the nicked GATC sequence and unbound the DNA from the mismatch creating an appropriate substrate for proper single-stranded exonucleases (An *et al*., 2000).

UvrD plays important role in the replication of small drug resistance plasmids by a rolling circle mechanism (Bruand *et al*., 2000) so, the viability of cells deficient of Rep (Petit *et al*., 2002) and DNA polymerase I (Moolenaar *et al*., 2000). The UvrD from the *E. coli* unbond both the double strand DNA and the DNA/RNA hybrids with the support of ATP of 3’ to 5’ polarity. As per successful unwinding cycle of the substrate, average of 4-5 base pairs is unwound (Matson *et al*., 1990).

UvrD DNA helicase protein is a concentration-dependent enzyme or called as stoichiometric. Despite of the length of the duplex fragment, the number of base pairs unbound is directly proportional to the UvrD concentration. The unwinding reaction proceeds in a 3' to 5' direction. But a single UvrD monomer can bind to the 3'- ssDNA or 3'- dsDNA junction of any DNA substrate is not capable to fully unwind even a short 18 bp duplex DNA since it requires two UvrD monomers that is capable to unwind short DNA substrates *in vitro* (Maluf *et al*., 2003).

Mutations in the UvrD gene result in multiple phenotypes, including increased sensitivity to UV, ionizing radiation, and alkylating agents, and increased rates of spontaneous and bromouracil-induced mutations. The mutants also show altered rates of genetic recombination and of precise transposon excision.

## 2.3 Recombinant Protein Expression in *E. coli*

Today, with the assistant of recombinant DNA technology and protein engineering, enzymes can be tailor-made to suit the requirements of the users or of the process. It is no longer necessary to depend on an enzyme's natural properties. The enzyme industry flourished in the 1980s and 1990s when microbial enzymes came onto the scene. Recombinant protein expression has altered dramatically all aspects of the biological sciences. The expansion in the number of proteins can be investigated both biochemically and structurally. The new commercial systems for recombinant protein expression with advanced protein purification techniques have made protein production prevalent throughout the biological and biomedical sciences. This advancement has enabled the research community to study thousands of low abundance and novel proteins from a large variety of organisms. According to the research, 31 recombinant proteins were approved for therapeutic use between 2003 and 2006; emphasize the importance of heterologous protein expression in biopharmaceutical research (Walsh, 2006). Apart from that, the development in multiple non bacterial recombinant expression systems are over the last three decades on yeast, baculovirus, and mammalian cell also been introduced.

*E. coli* is one of the prelemenary and most widely used hosts for the production of heterologous proteins (Terpe, 2006). *E. coli* genetics are far better understood than those of any other microorganism. It is used for massive production of many commercialized proteins. *Escherichia coli* are still the most preferable expression host for recombinant protein expression (Yin *et al*., 2007). The advantages in this expression system are, *E. coli* is easy to genetically manipulate, it is inexpensive to culture, and expression is fast, with proteins routinely produced in one day. Moreover, protocols for isotope-labelling for NMR spectroscopy and selenomethionine incorporation for X-ray crystallography are well established, making it highly suitable for structural studies. This expression system is excellent for functional expression of non-glycosylated proteins. Recently progresses in the major understanding of transcription, translation, and protein folding in *E. coli*, together with the availability of improved genetic tools. This made *E. coli* more important than for the expression of complex eukaryotic proteins. The genome factor of *E. coli* can be rapidly and precisely modified with ease, promoter control is not difficult, and plasmid copy number can be readily altered. This system also can modify the of metabolic carbon flow, avoidance of incorporation of amino acid analogise, formation of intracellular disulfide bonds, and reproducible performance with computer control. *E. coli* can accumulate recombinant proteins up to 80% of its dry weight and survives a variety of environmental conditions.

Despite that, it has many advantages and widespread use; there are also disadvantages to using *E. coli* as an expression host. The *E. coli* system has some setbacks, however, which has to be over rule for efficient expression of proteins. Proteins which are produced as inclusion bodies are often inactive, insoluble and require refolding. In addition, there is a problem producing proteins with many

disulfide bonds and refolding these proteins is extremely difficult. The transcription and translation are fast and tightly coupled. The rate enhancement often leads to a pool of partially folded, unfolded, or misfolded, insoluble proteins (Oberg *et al*., 1994). Thus, some targets, especially larger multi domain and membrane proteins, either fail to express in *E. coli* or express insolubly as inclusion bodies. Moreover, insolubility is not just restricted to heterologous proteins, as many bacterial proteins also cannot be produced in soluble form when over expressed in *E. coli (*Vincentelli *et al*., 2003). To improve the *E. coli* process as follows: (i) use of different promoters to regulate expression; (ii) use of different host strains; (iii) co-expression of chaperones and/or foldases; (iv) lowering of temperature; (v) secretion of proteins into the periplasmic space or into the medium; (vi) reducing the rate of protein synthesis; (vii) changing the growth medium; (viii) addition of a fusion partner; (ix) expression of a fragment of the protein; and (x) in vitro denaturation and refolding of the protein (Wong *et al*., 2008). Production can be improved with a promoter system like lac, tac, trc is been used. Promoter systems must be strong and tightly regulated so that they have a low-basal level of expression, easily transferable to other *E. coli* strains. Secretion of recombinant proteins by *E. coli* into the periplasm into the medium has many advantages over intracellular production as inclusion bodies. It helps downstream processing, folding and *in vivo* stability, and allows the production of soluble, active proteins at a reduced processing cost (Mergulhao *et al*., 2005).

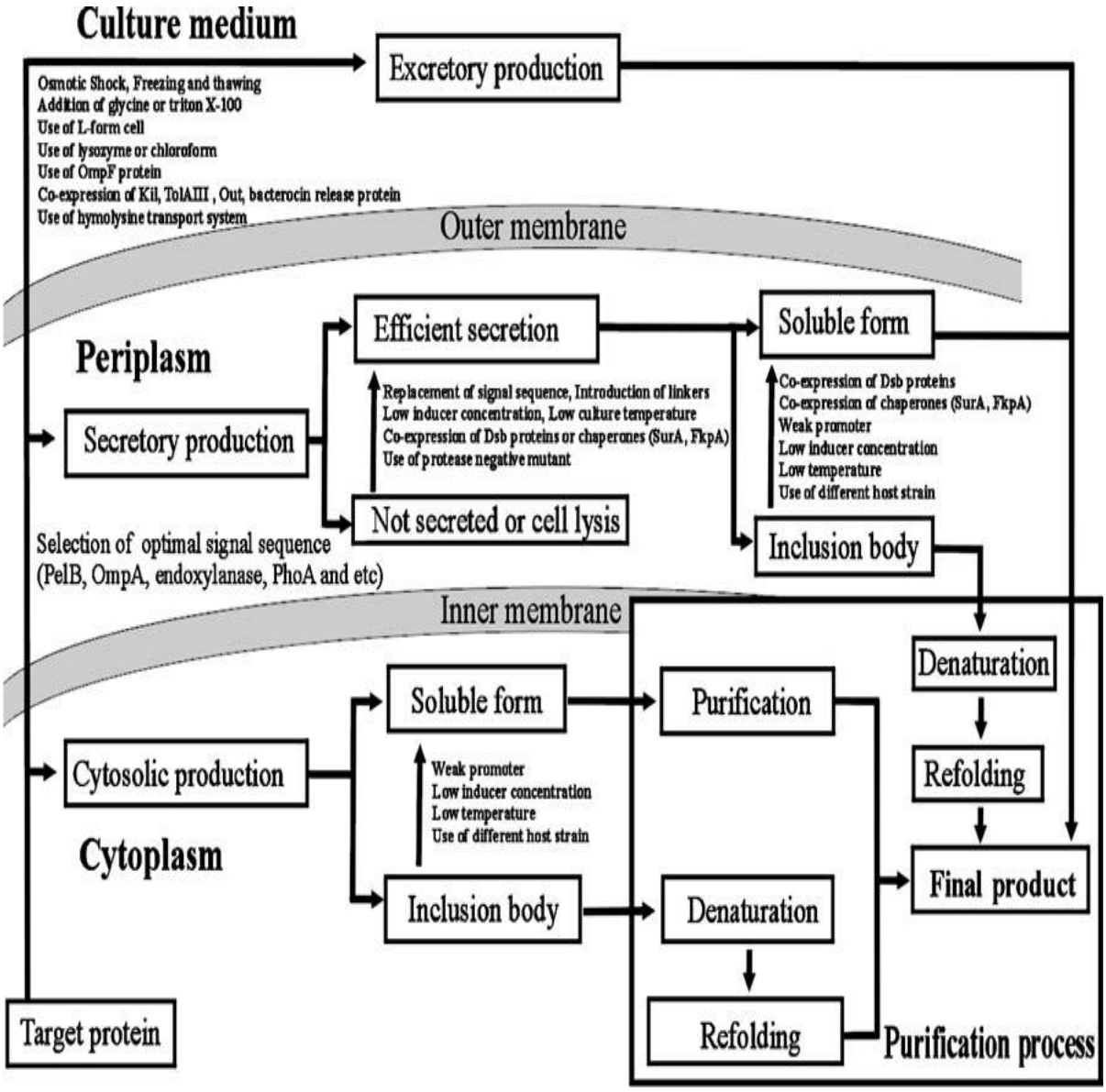



**Figure 2.1:** Strategies for the production of recombinant proteins in *E. coli*. Recombinant proteins can be produced in three compartments of *E. coli* in soluble or inclusion body forms, from which they follow different purification processes.

Source: Choi *et al.*, 2006

## 2.4 Downstream Process Challenges and Overcomes

The primary challenges in downstream processing of protiens include production yield, homogeneity and purity. Production of small amounts is typically straightforward. Biological activity enzyme molecules are closely related to their chemical and structural stability. Degradation may occur at various stages from production/purification through formulation, and storage. Two main categories of degradation are physical and chemical. Aggregation is the most common type of physical degradation and when it occurs the monomeric units bind to each other forming dimers, trimers, tetramers or even higher molecular weight aggregates size of which could range from nanometers to microns. It could be induced by various stresses such as temperature change, freeze/thaw, mechanical stress (agitation, pumping, and filtration, filling), and pH/conductivity change. Chemical degradations occur via oxidation, deamidation, isomerization, cross-linking, clipping and fragmentation (Goswami et al., 2013).

Modern biopharmaceuticals are commonly produced with defined media whose components are chemically characterized. In the past, yeast, meat, and soy extracts, produced by proteolytic degradation and extraction, were commonly used for the cultivation of bacteria and yeast cells. The standardization of such raw material was extremely difficult resulting in substantial batch to batch variations. So recently media for the industrial cultivation of bacteria are usually very simple and provide the essential sources of carbon, nitrogen and phosphate. Purity requirements for biopharmaceuticals vary depending on the particular application. Thus, it is not possible to specify absolute values. However, an important distinction can be made among the various impurities which are frequently categorized as critical or

non-critical. A non-critical impurity is an inert compound without biological relevance. On the other hand, endotoxins or growth factors secreted into the culture supernatant are examples of critical impurities, since they exert adverse biological activity. These impurities need to be traced throughout the process.

## 2. 5 Affinity Chromatography

Affinity chromatography was discovery 50 years ago (Cuatrecasas *et al.*, 1968). The sophisticated invention is to replace the traditional purification methods which conducted based on pH, ionic strength, or temperature. It has been progressively taking over 60% of all purification techniques by introducing affinity chromatography (Lowe, 1996). The vast applicable of this technique based on the fact that, any given bimolecular has to be purify usually consist of a recognition site where it can be bind by an ordinary or synthetic molecule.

The principle behind this mechanism is that targeted proteins were separated using affinity chromatography which relies on the reversible interactions between the purified protein and the affinity ligand conjugate to chromatographic matrix. Most of the proteins have an intrinsic recognition site that serves to choose the proper affinity ligand. In this case, the main factor to consider is the binding between target protein and the affinity ligand must be specific and reversible (Zachariou, 2008). Recombinant tagged proteins purification of proteins can be effortless and straightforward technique if the protein of interest is tagged with a commonly known sequence referred to as a tag.

His-tag can attach to either the N- or C-terminus site. Purification efficiency is achieved optimal binding during the freely accessibility of his-tag to metal ion support (Dong *et al*., 2010). Histidine tags consist of strong affinity for metal ions such as Co2+, Ni2+, Cu2+, and Zn2+. Basically there are two metal binding support materials used for protein purification which is Iminodiacetic acid (IDA) and nitrilotriacetic acid (NTA). But the disadvantage in iminodiacetic acid is its latent for metal ion leaching and eventually leads to a lower protein yield. This is due to Iminodiacetic acid (IDA) matrix to chelate transition metals carries three coordination sites only. IDA matrices are weakly bound to the metal ion with three-coordinate matrix. Metal leaching from the matrix during purification causes decreased yields and impure products (Porath *et al*., 1975).

In modern technique support materials, including nickel-nitrilotriacetic acid (Ni-NTA) and cobalt-carboxymethylasparate (Co-CMA) shown on the Figure 2.2, this shows limited leaching and, resulted in more efficient protein purifications. When the His-tagged protein is bind to the immobilized chelating agent, it can be eluted by exposing to a competing agent for the chelating group (imidazole) or an additional metal chelating agent (EDTA). The benefits of using His-tags for protein purification are because of the small size of the affinity ligand. The indirectly reduces effects on the folding of the protein. The placement of His-tag on the N-terminal end of the protein can easily be removed using an endoprotease.

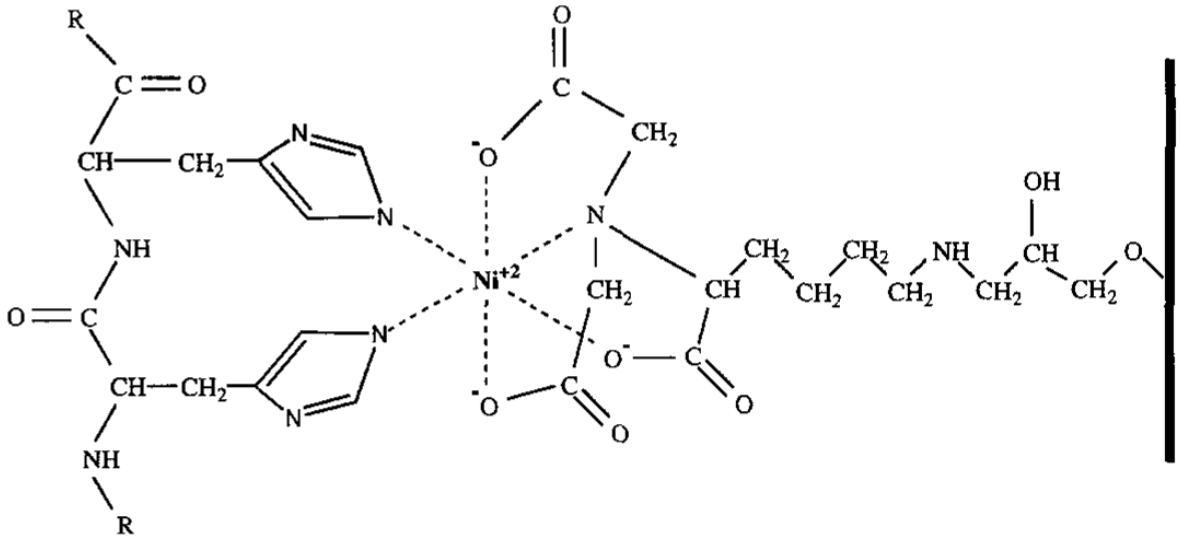
The advantage of using the polyhistidine tags for purification, His-tag can bind to proteins under both native and denaturing conditions. Denaturing conditions of protein becomes important when target proteins are excrete in inclusion bodies. So denatured protein can be solubilised but later has to refold to show the functionality. The disadvantages of using his-tag protein purification listed as potential degradation

of the His-tag, dimer and tetramer formation, and co elution of other histidine-containing proteins. If the histidine residues are proteolytically degraded, the affinity of the tagged protein is highly decreased leads to a lower protein yield. A protein has a His-tag added to its structure but has the potential to form dimers and tetramers in the presence of metal ions. This can lead to erroneous molecular mass estimates of the tagged protein. Lastly using His-tags is coelution of proteins that naturally have two or more adjacent histidine residues.

A disadvantage of polyhistidine affinity tags could be nonspecific binding of untagged proteins even though histidine occurs relatively randomly about 2% of total protein residues but some cellular proteins exhibit two or more adjacent histidine residues. These proteins have an affinity towards the IMAC matrix and may coelute with the protein of interest, resulting in significant contamination of the final product. These problems are more prominent in mammalian systems then in *E. coli* systems. Disulfide bond formation between the proteins of interest with other proteins can also lead to contamination. Introducing 2-mercaptoethanol into the loading, washing, and elution buffers overcome this problem. Nonspecific hydrophobic interactions can also cause some copurification with the desired protein.

The problem can be overcome by place the tag to specifically protein at optimal position. There are some potential problem like inaccessibility of the protein tag to the immobilized metal because of occlusion of the tag in the folded protein can be rectify by positioning the affinity tag to the opposite terminal of the protein or conduct the purification under denaturing conditions. Polyhistidine affinity tags are small size can be integrated easily into any expression vector (Crowe *et al*., 1994).

a. Nickel-nitriloacetic acid ($Ni^{+2}$-NTA)




b. Cobalt-carboxylmethylaspartate ($Co^{+2}$-CMA)

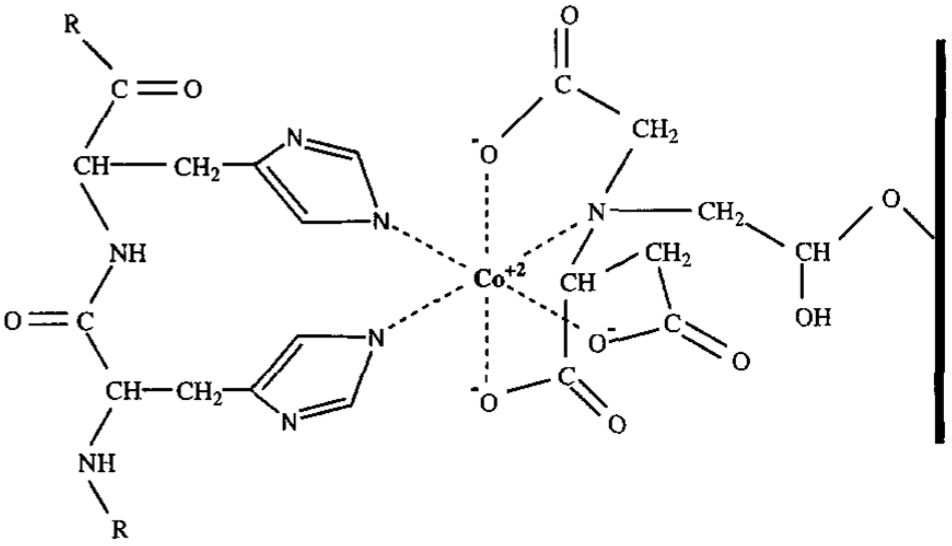



**Figure 2.2:** Models of the interactions between the polyhistidine affinity tag and two immobilized metal affinity chromatography matrices, (a) The nickel–nitrilotriacetic acid matrix ($Ni^{+2}$–NTA).

Source: Joshua *et al*., 2000

.

## 2.6 Western Blot Analysis

The western blot is a technique commonly used to identify, quantify, and determine the size of specific proteins. This is fusion method derived from Southern blotting, where it developed to identify specific DNA sequences among DNA fragments separated by gel electrophoresis, and northern blotting, used to identify and quantify RNA and to verify size. This process too involves gel electrophoresis to separate RNA. Studies by Towbin et al.,1979 during late 70's uses electrophoretical method to separate protein with the uses of polyacrylamide–urea gels. Followed by, the transferring method onto a nitrocellulose membrane. Later the sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE) becomes more famous by discovery of Burnette (1981). These progresses eventually create this method called western blotting. It's known as protein blotting or immunoblotting device for detect proteins. Gel electrophoresis method used to separate native or denatured proteins. The proteins are then transferred to a membrane for exposure using antibodies specific to the target protein. The technique evolved, undergoes many troubleshooting and improving the technique (Kurien and Scofield, 2009).

Proteins separated using gel electrophoresis has certain separation factor to be considered such as isoelectric point, molecular weight, electric charge, or all. Sodium dodecyl sulfat (SDS) is most common electrophoresis separation uses polyacrylamide gels and running buffers. Proteins contain various electrical charges, but when the protein mixture treated with SDS, they turn out to become denatured and negatively charged. This allows separation of proteins by molecular weight. The reducing agent in the mixture eliminates disulfide bonds and boiling of samples can facilitate denaturing.

In continuation from protein separation in complete electrophoresis, the proteins can be transferred from the gel onto a membrane made of nitrocellulose, polyvinylidene difluoride (PVDF), activated paper, or activated nylon (Kurien and Scofield, 2006). Procedure for transferring proteins from a gel to a membrane is called as Electroblotting. This process uses an electrical energy to pull proteins from the gel onto the membrane. The transferred has a special method known as gel-membrane sandwich (wet transfer) carried out by immersion of a membrane been soaked in transfer buffer (semidry transfer). The effectiveness of protein transfer is dependent on the type of gel used, the molecular mass of the protein, and the type of membrane.

Precaution has to taken during antibody introduction, this is important to prevent interactions between the membrane and the antibody chosen to detect the target protein. Blocking of nonspecific binding can be achieved by placing the membrane into a dilute solution of protein such as bovine serum albumin and nonfat dry milk. The blocking buffer should be suitable for the specific antiserum and for the type of membrane. Blocking helps protecting from any potential nonspecific binding sites on the membrane, thus reducing background "noise" in the final product of the western blot, eliminating false positives and providing a clear result.

The membrane further incubates with primary antibody, wash, reblock, and then incubate with secondary antibody and repeat the wash again. Determination of the optimal concentration of antibodies is vital before running all the samples. It's a prime determinant of the sensitivity of the assay (Burnette, 1981). The antibody concentration should be optimized to provide the best signal to noise ratio. Monoclonal and polyclonal antibodies can be used for western analyses, with advantages and disadvantages in using either type (MacPhee, 2010).

The probes that are labeled and bound to the protein of interest detected on the western blot. There are number of detection methods are available such as, colorimetric, radioactive, and fluorescent methods. But, chemiluminescent detection is used most often used. This intensify the chemiluminescence (ECL) a sensitive method used for relative quantitation of the target protein (Kurien and Scofield, 2006; MacPhee, 2010). The primary antibody binds to the protein of interest and follows by the secondary antibody, usually linked to horseradish peroxidase, is used to cleave a chemiluminescent agent. The reaction product produces luminescence, which is related to the amount of protein. Only a single light detector is required, and the light is detected by photographic film or by a charged-couple device camera.

Even though the procedure for western blot is simple, many problems can arise, leading to unexpected results. The problem can be grouped into five categories: (1) unusual or unexpected bands, (2) no bands, (3) faint bands or weak signal, (4) high background on the blot, and (5) patchy or uneven spots on the blot.Unusual or unexpected bands can be due to protease degradation, which produces bands at unexpected positions. In this case it is advisable to use a fresh sample which had been kept on ice or alter the antibody. If the protein seems to be in too high of a position, then reheating the sample can help to break the quaternary protein structure. Similarly, blurry bands are often caused by high voltage or air bubbles present during transfer. In this case, it should be ensured that the gel is run at a lower voltage, and that the transfer sandwich is prepared properly. In addition, changing the running buffer can also help the problem. Nonflat bands can be the result of too fast of a travel through the gel, due to low resistance. To fix this the gel should be optimized to fit the sample. Finally, white (negative) bands on the film are due to too much protein or antibody.

## 2.7 MutL Gene

MutL gene also known as Methylmalonyl Co-A mutase was first discovered in 1955 on rat liver and sheep kidney. The gene encoding for this enzyme in humans is knownas MUT, which corresponds to chromosome 6p12-21.2. The enzyme commission number of this MutL gene is 5.4.99.2 and it is grouped into mutase that catalyzes the shifting of a functional group from one position to another within the same molecule.

### 2.7.1 Function of MutL

The main function of MutL protein is its involvement in the repair of mismatches in DNA. This protein is required for dam-dependent methyl-directed DNA mismatch repair. In Sancar and Hearst study suggested MutL also known as a "molecular matchmaker", where a protein that promotes the configuration of a established complex between two or more DNA-binding proteins in an ATP-dependent manner without taking part of the final effectors complex. The ATPase activity of MutL is stimulated by DNA. The biochemistry of MutL action in *mutHLS* repair action is well studied than that of MutS (Modrich *et al*., 1996).

### 2.7.2 MutL is a Molecular Switch in DNA Mismatch Repair

Proteins that hydrolyze NTP are generally categorized into two classes which are motor proteins like myosin, actin or DNA helicase or signaling proteins like G-proteins. The interesting structural and functional comparison between the two classes has been made and guides characterization of MutL (Vale, 1996). MutL does not characterize under a motor protein for the following reasons. MutL hydrolyzes ATP about 100-fold more slowly than any known motor protein (Ban and Yang, 1998). Depsite it binds to DNA, in return activates its ATPase activity, but MutL does not possess a helicase activity. The function in mediating interactions between MutS and MutH, suggests that MutL is an ATP-operated signaling molecule. Certain criterias need to be fulfilled for the ATPase activity to be utilized by MutL for signaling. The primary step, different states of the nucleotide–MutL complex must be used to communicate with different partners. By activating MutH, MutL is known to interact with the DNA helicase UvrD and initiate a role in recruiting DNA polymerase III to complete the repair process after MutH nicks the daughter strand (Hall *et al.*, 1998).

Secondly, the ATPase cycle of MutL must be regulated, structural species of MutL is dominant at a particular time and serves a specific function. DNA single-stranded in particular, stimulates the ATPase activity of MutL. The analyses of kinetic effects of ssDNA on the MutL ATPase activity revealed that both Km and Kcat are increased and ssDNA lowers the affinity of MutL for ATP while enhancing the rate of ATP hydrolysis. During the mismatch repair process, ssDNA occurs after MutH nicks bind daughter strand and exonuclease starts to remove nucleotides and

exposes the template DNA. At that stage, MutH is no longer needed, while helicase and DNA polymerase take the control in reaction. MutL serves as a molecular switch that recruits different proteins at various steps in the mismatch repair process with control of ATP. Apart from it, MutS, MutL and their homologues involve in repair of oxidative or carcinogen damaged DNA and the connections between mismatch repair and programmed cell death and requirement of a molecular switch or adapter to coordinate various DNA repair pathways and to choose repair versus apoptosis (Zhang *et al.,* 1999).

### 2.7.3 MutL in Eukaryotes

In eukaryotes, MutL homologs are called as MLH and PMS. The *pms* gene was named for the phenotype of postmeiotic segregation before it was cloned and found to be homologous to MutL (Kramer *et al.*, 1989). As in *E. coli* proteins, eukaryotic MutS homologs, MLHs, and PMSs play an essential role in DNA mismatch repair. Mutations of *MLH1* have been identified in human leukemia and lymphoma cell lines as well (Hangaishi *et al.,* 1997). Eventhough the linkage between cancer susceptibility and DNA mismatch repair has been established, the relationship between DNA repair proteins and tissue-specific development of cancer is unclear. MutL homologs in eukaryotes, present in humans, are clearly essential for normal cell growth.

All members of the MutL family have a conserved region of ~300 residues at their terminus and a rather diverse C-terminal region of 300 to 500 residues. They often form homo- or heterodimers in solution. The majority of the reported mutations with dominant mutator phenotypes in *E. Coli* MutL (Aronshtam and Marinus 1996)

are within the conserved N-terminal region. *E. coli* MutL has been proposed to mediate the interaction between the endonuclease MutH and the mismatch recognition protein MutS. MutS and MutL were shown to be linked in heteroduplex loop formation by electron microscopy (Allen *et al*., 1997). Addition of MutL increase the region protected by MutS in a DNA foot printing analysis (Lahue *et al*., 1989). Modification in the DNA footprint could result from interactions between MutS and MutL. Evidence of a physical interaction between MutL and UvrD has been reported (Hall *et al*., 1998). However, MutL interacts with MutH and MutL switches from activating MutH to recruitment of the helicase to a repair site MutL and its homologs have not attracted as much scrutiny as the MutS family even though mutations in the MutL and MutS families display similar phenotypic defects. The main reason for the shortage of attention to the MutL family is due to a lack of an identified activity of the MutL family.

### 2.7.4 Interactions between MutL and UvrD

MutL gene is needed to stimulate the UvrD helicase to unwind duble strand DNA with a blunt end. But stimulation of UvrD on a nicked circular DNA substrate requires a MutS–MutL-mismatch complex (Dao and Modrich, 1998; Yamaguchi *et al*., 1998). The MutL gene was identified to consist of 218 residues (residues 398–615) in C-terminal to interact with UvrD according to two-hybrid analyses and pull-down assays (Hall *et al*., 1998). The same approach has taken reveal that the same C-terminal region of MutL interacted with MutH (Hall and Matson, 1999). The MutL C-terminal plays a major role in identification of homo- or hetero-dimerization of MutL and its homologs (Wu *et al*., 2003).

Moreover, the amino-acid sequence of the C-terminal region is specifically various among MutL homologs. MutL revitalizing the 3’ -5’ helicase activity of UvrD helicase and, in the existence of a mismatch and MutS, MutL loads UvrD from a nick onto either the nicked or continuous strand, based on the nick is 5’ or 3’ to the mismatch site, so that UvrD unbinds DNA strands toward the mismatch (Dao and Modrich, 1998; Yamaguchi *et al.*, 1998). The interaction between MutL and UvrD are called to direct, as recommended by previous studies (Hall *et al.*, 1998; Spampinato and Modrich, 2000) or change DNA structure to amplify the helicase activity. Figure 2.3 is a schematic diagram of intraction between MutL protein and UvrD helicase involved in the methyl mismatch repair in *E. coli*.