CHAPTER 1

Introduction
Mycobacterium tuberculosis is an obligate mammalian pathogen that adapts to host challenges during the course of infection and can enter into a latent/persistent stage in situ for years. The regulation of gene expression plays an important role in different stages of infection (Raman et al, 2004; Akif et al, 2006 & Manabe et al, 2000). The pathogen is responsible for the cause of ~3 million deaths every year (World Health Organization, 2006). Tuberculosis infection can be broadly divided into three stages: (1) establishment of infection and acquisition of cell-mediated immunity, (2) a chronic, latent phase and, in some individuals, (3) a reactivation phase that is characterized by rampant bacterial growth and the symptoms of tuberculosis (Murray, 1999).

It can persist for a long periods in a dormant or persistence state resulting in asymptomatic chronic infections, which may become active in replicating bacilli after several years when the host becomes immuno-compromised (Urlichs et al, 2006). These bacilli have the ability to remain dormant until the defence system of host is compromised, as is the case in HIV infections. With the increase in number of HIV infected patents, the number of people infected with both tuberculosis and HIV represent a potential danger as increasing number of actively infected host are capable of transmitting infection has been noted (Manabe et al, 2000). Simultaneously, increase in the emergence of multi-drug resistance tuberculosis (MDR-TB), which poses a significant threat to the control of tuberculosis (Zhang, 2004). The current tuberculosis chemotherapy results in killing growing bacilli but is largely ineffective in destroying persistent or dormant bacilli, leading to prolonged therapy. An ideal chemotherapy regimen for tuberculosis that would, (1) reduce the duration of treatment, (2) be active against drug resistance strains and (3) be capable of eradicating latent tuberculosis infections. Therefore a comprehensive understanding of the cellular process occurring during non-replicating persistence should facilitate attempts to interfere with this process in a rational manner (Cho et al, 2006).

During the several stages of infection, M. tuberculosis encounters a changing host environment, in response to which the bacillus must activate defence and repair mechanisms and reprogram its physiology to ensure survival. M. tuberculosis exhibits significant changes in gene expression during the latent or persistence stage of infection. Proteomic analysis of the nutrient starved bacteria has suggested an induction of stringent response and genes that may play a role in long-term survival
within the host. The large number of putative transcription regulators identified in the
*M. tuberculosis* genome sequence indicates that much of the regulation required for
these adaptations by *M. tuberculosis* occurs at the level of transcription. Regulators
like Rv3291c; a member of the Lrp/AsnC family of transcriptional regulators has been
reported to be up-regulated over 15-folds in the nutrition starvation model of
persistence and is believed to be important for maintaining the long-term survival of
the pathogen (Betts *et al.*, 2002).

### 1.1 Lrp/AsnC Family of Transcriptional Regulators

The Lrp/AsnC family of transcriptional regulators includes proteins that act as
global or specific regulators of transcription have been isolated from many
prokaryotes, including both bacteria and archaea (Brinkman *et al.*, 2000 & Kyfrides
*et al.*, 1995). These proteins are known to regulate amino-acid metabolism and related
processes. Members of this family have been identified in 45% of bacterial genomes
and 95% of archaea. To date, there are no confirmed homologues available in
eukaryal genomes, indicating that this family is probably restricted to prokaryotes
(Brinkman *et al.*, 2003). Although one of the recent reviews (Kawashima *et al.*, 2008)
has traced back the origin of eukaryotic modulations of transcription to archaea on the
basis of some geometry of the three α helices in DBDs (DNA binding domains) to the
homeodomains of eukaryotic transcriptional regulators.

The family is named after two proteins from *Escherichia coli* implicated in the
control of amino acid metabolism; the Lrp (Leucine responsive regulatory protein)
and AsnC (regulator of asparagine synthase C gene product), which in addition to
sharing ~25% sequence identity, stimulate operon (Lrp - *ilvIhi* and AsnC – *asnA*)
expression and a pathway related end product prevents this stimulatory effect (Willins

Lrp is a global regulator that controls the expression of a large number of
operons in *E. coli*, including those involved in the synthesis and degradation of amino
acids (Brinkman *et al.*, 2003). AsnC is a specific regulator of the *asnA* gene, which
codes for asparagine synthetase, responsible for converting aspartate to asparagine in
an ATP-dependent reaction, and an autorepressor of its own expression (Kolling *et al*.,
1985). As the name implies, many Lrp-responsive operons are co-regulated by L-
leucine that can either antagonize or potentiate the effects of Lrp (Calvo *et al.*, 1994).
AsnC, on the other hand appears to be responsible for controlling asparagine levels via negative feedback regulation of asparagine synthetase expression (Kolling et al., 1985 & de Wind et al., 1985).

Members of the Lrp/AsnC family typically have a molecular mass of ~15 kDa but populate a range of multimeric species in solution that include dimers, tetramers, octamers and hexa decamers (Willins et al., 1991; Madhusudhan et al., 1995; Brinkman et al., 2000; Jafri et al., 1999 & Chen et al., 2001).

Homologous of Lrp and AsnC are referred to as the Feast/Famine Regulatory Proteins (FFRPs). Calvo and Matthews coined the term Feast/famine Regulatory Proteins (Calvo et al., 1994) for Lrp/AsnC proteins to summarize the general function of these proteins.

1.2 Feast/Famine regulatory proteins (FFRP)

The term Feast/famine is used for a class of proteins that are involved in sensing the concentration of amino acids/effector molecules from the nutrient media and effecting appropriate changes to gene expression (Suzuki, 2003). The best-characterized bacterial member of this family is the E. coli Lrp. On sensing a high concentration of leucine, E. coli Lrp shifts its metabolism to a more heterotrophic mode, activates absorption of nutrients, accelerate cell replication, and changes the pathogenicity of the organism. For this adaptation, Lrp regulates transcription of a number of genes in various ways (Calvo et al., 1994 & Newman et al., 1995) where as depending upon concentration of asparagine; AsnC down-regulates biosynthesis of this amino acid (Kolling et al., 1985). The feast or famine type of change may not be the only survival mechanism in an organism that utilizes response by FFRPs. For example, unlike many other archaea, T. volcanium survives aerobic and anaerobic environments, and it is likely that FFRPs are involved in this adaptation (Suzuki, 2003).

Insights into the relationship between the structure of Lrp/AsnC family proteins and their function have come from X-ray crystallography. To date, six X-ray crystal structures (other than the present work) from different sources in this family have been reported, two from archael source; LrpA (PDB ID 1H1G) from Pyrococcus furiosus (Figure 1.2a, Leonard et al., 2001) and Pyrococcus sp. OT3 FL11 (PDB ID 1RI7) (Figure 1.2b, Koike et al., 2004) and also in complex with DNA (PDB ID
2E1C) (Figure 1.2c, Yokoyama et al., 2007) and from bacterial source E. coli Lrp (PDB ID 2GQQ) (de Los Rios et al., 2007) and AsnC in complex with asparagine (PDB ID 2CG4) (Figure 1.3a, Thaw et al., 2006), Neisseria Meningitides NMB0573 complexed to leucine and methionine, respectively (PDB ID 2P5V, 2P6S and 2P6T) (Ren et al., 2007) and LrpC from Bacillus subtilis (PDB ID 2CFX) (Thaw et al., 2006).

On the basis of domain organization FFRPs can be divided into four types that are described below.

1.2.1 Types of FFRPs on the basis of domain composition:

Many FFRPs, including the three E. coli proteins, have ~160 amino acid residues, each forming an N-terminal DNA-binding domain (DBD) and a C-terminal assembly domain combinedly forming a full-length FFRP (FL-FFRP). Some other eubacterial and archaeal FFRPs have only ~80 amino acid residues, corresponding to the C-terminal halves of FL-FFRPs, i.e., demi (DM)-FFRPs. These FFRPs do not bind DNA, unlike their counterpart (Koike, et al., 2004). In the genome of Pyrococcus OT3, 11 genes code for FL-FFRPs three for DM-FFRPs (Yokoyama et al., 2006). The protein Bs0711124 from the eubacterium Bacillus subtilis (Suzuki et al., 2003a) and archaeal proteins including Lrs14 (Sedelnikova et al., 2001) from Sulfolobus solfataricus correspond to the N-terminal half of an FL FFRP: N-DM-FFRPs (Suzuki et al., 2003a). Some other open-reading frames of archaeal source code for proteins, which has a domain similar to the DBD of an FL-FFRP, but the rest of the protein is longer and thus different from the C domain (Suzuki et al., 2003a). Figure 1.1 contains schematic presentation for different types of FFRPs:-
Figure 1.1 Types of FFRPs on the basis of domain organizations: (a) A full-length FFRP (FL-FFRP) is composed of the N terminal DNA-binding and C terminal assembling domains respectively. (b) A demi-FFRP (DM-FFRP) corresponds to the C-terminal half only of an FL-FFRP. (c) An N-demi-FFRP corresponds to the N-terminal half of an FL-FFRP, followed by some additional residues. (d) Some proteins have DNA-binding domains of the FFRP type at their N termini, followed by amino acid residues possibly forming non-FFRP types of assembling domains.

Different types of FFRPs in both archaeal and eubacterial sources are identified and are described below:

**FFRPs from Archaeal Sources**

The presence of homologous of *E. coli* Lrp and AsnC in archaea was first reported by (Kyrfrides et al., 1995). Since then many FFRPs have been identified using archaeal genomic sequences and has been found to be widely distributed in them (Figure 1.2, structurally characterized members of archaeal FFRPs).

Kawashima and coworkers in 2008 have shown that genome of *Thermoplasma volcanium* contains six FFRPs; five belong to FL-FFRP and one to DM-FFRP type respectively. TvfFL1 was shown to be important for the aerobic/aerobic adaptation and synthesized exclusively in aerobic conditions.
Figure 1.2 Structurally characterized members of archaeal FFRPs (a) Dimer of LrpA (111G) from *Pyrococcus furiosus* (b & c) dimer of FL11 from *Pyrococcus sp. OT3* (1R17) and in complex with DNA (2EIC) (d) TTH0845 from *Thermus thermophilus* (2DJW) (e) DM1 from *Pyrococcus sp. OT3* (2EIA) (Figure of DM1 adapted from Yokoyama *et al.*, 2007).
Production of TvFL2 was more in aerobic conditions whereas TvFL5 in anaerobic conditions. TvFL2, TvFL3 and TVFL5 all the three bind to the promoter of Rieske ion sulfur proteins and nadh dehydrogenase. Yokoyama et al., 2006 identified various types of FFRPs in archaeal genome are listed below.

Table 1.1 Numbers of three type of FFRPs identified in Archaeal genome (Yokoyama et al., 2006)

<table>
<thead>
<tr>
<th>Archaeal Source</th>
<th>FL-FFRP</th>
<th>Demi-FFRP</th>
<th>N-Demi FFRP</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sulfolobus solfataricus</em></td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td><em>Pyrococcus sp. OT3</em></td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td><em>Pyrococcus furiosus</em></td>
<td>10</td>
<td>3</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td><em>Pyrococcus abyssi</em></td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td><em>Archaeoglobus fulgidus</em></td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td><em>Halobacterium sp. NRC-1</em></td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td><em>Thermoplasma volcanium</em></td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td><em>Thermoplasma acidophilum</em></td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><em>Aeropyrum pernix</em></td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td><em>Methanococcus jannaschii</em></td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><em>Methanobacterium</em></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>thermoautotrophicum</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FFRPs from Bacterial Sources**

In the genome of *E. coli*, three FFRPs are coded: Lrp, AsnC and YbaO. Anderson and co-workers first identified the Lrp gene in 1976 as a locus (livR) affecting the transport of branched-chain amino acids and its regulation (Calvo et al., 1994). Lrp is the best-characterized member of all FFRPs in biological terms. AsnC is a non-orthologous homolog of Lrp; depending upon concentration of asparagine i.e. with increasing concentration of asparagine, AsnC down regulates autotrophic biosynthesis of this amino acid (de Wind et al., 1985, Kolling et al., 1985). The function of the third *E. coli* FFRP, YbaO, is unknown, although its amino acid
sequence resembles that of another FFRP, Grp (the glutamate uptake regulatory protein) (Zeng et al., 1997), which regulates uptake of glutamate in another eubacterium (Peekhaus et al., 1995).

In the genome of Neisseria sp. two proteins belonging to the Lrp/AsnC families have been reported, genome of Neisseria meningitidis serogroups A and B and Neisseria gonorrhoeae have been sequenced and each shown to contain two Lrp/AsnC family regulators (Tettelin et al., 2000). The two regulators exemplified by the gene products of NMB0573 (annotated asnC) and NMB1650 (annotated lrp) in N. meningitidis serotype B are each highly conserved between the two species of pathogen (≥99% sequence identity) and are 27% identical to each other (Ren et al., 2007).

Bacillus subtilis has six FL-FFRPs and an N-demi FFRP, which corresponds to the N-terminal DBD of an FL-FFRP (Suzuki et al., 2003a). bsLrpC (bsFL3) regulates bslrpC gene, and has been predicted to be involved in sporulation (Beloin et al., 2000). bsLrpA (bsFL4) and bsLrpB (bsFL5) are involved in the processes regulated by KinB (membrane bound kinase regulate the process of sporulation in B. subtilis). AzlB (bsFL6) regulates the azlBCD operon, which is involved in transport of branched chain amino acids, isoleucine, valine and leucine (Belitsky et al., 1997). Figure 1.3 contains the structure of E. coli AsnC showing the basic domain organization of FFRP and structural divergence among the members of structurally characterized members of bacterial FFRPs.
Figure 1.3 (a) Structure of *E. coli* AsnC with bound asparagine (2CG4) (b) Superimposition of bacterial FFRPs, *E. coli* AsnC (2CG4) – Wheat; *E. coli* Lrp (2GQQ) – Light Pink; *Bacillus subtilis* LrpC (2CFX) – Light Green and *Neisseria Meningitides* NMB0573 (2P5V) – Yellow.
1.2.2 Modular Structure of FFRPs

On the basis of structurally characterized FL-FFRPs and genetic experiments done by Platko and co-workers, 1993, the basic modular architecture has been identified as an arrangement of two domains, N-terminal domain contains sites of DNA binding (Figure 1.3a) and C-terminal domain is responsible for higher oligomeric association (Figure 1.3b). DM-FFRP on the other hand have only C-terminal oligomerization domain.

Structurally characterized members from this group include two structures from archaeal source and four structures from bacterial sources as describe earlier. The basic subunit of protein is dimer which leads to higher order oligomer. Members of FFRPs are characterized by its low sequence identity below 30% with same overall structure (Yokoyama et al., 2007). The two basic domains of FFRPs are:-

1.2.2.1 DNA binding Domain

LrpA from *P. furiosus* was the first Lrp-like protein whose three-dimensional structure has been solved (Leonard et al., 2001). The structure revealed that the N-terminal part of the protein consists of a helix-turn-helix (HTH) domain, a fold generally involved in DNA binding. The N-terminal HTH has previously been predicted for several Lrp-like proteins by computer-based analysis of the primary amino acid sequence, as well as by the observation that mutations in this region affect DNA-binding in several Lrp-like proteins (Platko et al., 1993; Enoru-Eta et al., 2000 & Ouhammouch et al., 2000)

Sequence alignment studies done with different FFRPs show conservation of particular type of amino acid, i.e. hydrophobic residues, most importantly Val, Leu, Ile and Met, positioned in the 3.6 periodicity in α helices, Gly in particular types of loops and turns are some specific characteristics of DNA binding domain (Suzuki et al., 2003). The structure of the HTH domain of LrpA closely resembles to the other prokaryotic regulators like CRP (McKay et al., 1981) which is actually a winged-HTH, with a two-stranded β-hairpin flanking the HTH, the spatial arrangement of all three helices is very similar to that in the LrpA HTH. The LRP-HTH also structurally resembles to the HTHs of other transcriptional regulators like BirA and LexA (Brinkman et al., 2003).
Chapter 1

Architecture of the DNA-binding domain. The DNA-binding domain is composed of three α helices (α1, α2, and α3), helix-turn-helix motif formed by α2 and α3 plus one more helix α1 shielding an inner hydrophobic core. Such type of domain is not specified to FFRPs, but is found in a large number of proteins (Suzuki et al., 1995 & Suzuki et al., 2003a). One of the α helices (in most of the cases the third helix) protrudes from the rest of the domain, forming a convex surface, which is complementary to the concave surface of the DNA major groove. By best fitting into the DNA major groove, α-helix 3 (i.e. the recognition helix) approaches the bottom, thereby reading the nucleotide sequence via chemical interactions between the amino acid side chains and the chemical groups of the basepairs.

Among the predicted set of three α-helices the content of Lys and Arg was found to be highest in α-helices 3; the two types of residues often used to fix the geometry of DNA-recognition helices relative to the DNAs by binding to phosphate groups in the backbones indicating that the third helix is designed for DNA recognition. (Suzuki, 1993, 1994 & Suzuki et al, 1994)

Mode of DNA Recognition. Sequences of identified binding sites for Lrp-like proteins in distinct prokaryotic promoters often lack perfect inverted repeat elements. Compared to the almost perfect inverted repeats of binding sites of other global regulators such as CRP, fumarate and nitrate reduction regulator (FNR) (Barber et al., 1993), binding sites of Lrp-like proteins cannot easily be distinguished in target promoters (Calvo et al., 1994). Systematic evolution of ligands by exponential enrichment (SELEX) has been used to reveal the optimal binding sequences of different Lrp homologues and contain sequence elements with dyad symmetry and require specific bases at specific positions (Brinkman et al., 2000). Consensus sequence of 15 bp has been recognized by SELEX showing palindrome in part, YAGHAWATTWTDCTR, where Y = C or T, H = not G, W = A or T, D = not C, and R = A or G, contains clear dyad symmetry and is very similar to the one defined earlier (Cui et al., 1995). Target promoter contain a number of binding sites and to which binding is usually cooperative (Madhusudhan et al., 1995)

E. coli Lrp is the only FFRP whose target genes have been identified by extensive genetic and biochemical experiments. Its DNA-binding specificity has been best characterized by SELEX (Cui et al., 1995) and footprinting experiments (Wang et al., 1993a, b & Marasco et al., 1994). Summarizing all these results, the ideal
binding site of an Lrp dimer is deduced as $\text{AGAATTTTATTCT}$, where five bases each at the two ends, complementary to each other, sandwich a run of three T bases, or a run of three A bases along the complementary strand (Suzuki, 2003a & 2003b). For another *E.coli* FFRP, AsnC; four 13-bp sequences in asnA - asnC promoter were identified, separated by a regular insertion of 18 bps each (Suzuki, 2003a & 2003b).

Only FFRP that has been crystallized with DNA, FL11 binds to 13 bp stretch of a DNA duplex $\text{TGAAAWWTTTCA}$, where $W$ is T or A (Yokoyama *et al.*, 2005 & 2007). Changing any base in $\text{TGAAAWWTTTCA}$ weakens the interaction (Yokoyama *et al.*, 2007). Consensus binding sequences of *E. coli* Lrp and various other FFRPs are summarized into the form $\text{NANBNCNDNEWWWNENnNcNBNA}$, where e.g., $\text{NA}$ is a base complementary to $\text{NA}$ (Cui *et al.*, 1995 & Suzuki, 2003)

**Mode of DNA binding** A Pair of $\alpha$ helices 3 in an FFRP dimer binds to 5 bp each at the two ends facing the major groove of the DNA (Suzuki *et al.*, 2003 & Yokoyama *et al.*, 2005). The separation between pairs of helices 3 in the FFRP crystal structures was found to be even shorter than the separation in the dimer of catabolite activator protein (CAP) (Schulz *et al.*, 1991), which is known as one of the shortest of such.

**Bending caused by binding** It has been observed in case of Lrp, LrpA, LrpC and PutR binds to DNA and induces bending (Brinkman, *et al.*, 2000; Jafri, *et al.*, 1999; Wang *et al.*, 1993b; Beloin *et al.*, 2000 & Tapias *et al.*, 2000). Recently available crystal structure of FL11 in complex with DNA also shows a bending of $55^\circ$, when analysed with without DNA bound structure. The two DNA binding domain opened relative to each other, but the opening is not enough to keep the DNA straight (Yokoyama *et al.*, 2007). *E. coli* Lrp induces a bend of $52^\circ$ upon binding to single site and the angle of bending was increased to at least $135^\circ$ when it bound to two adjacent sites (Wang *et al.*, 1993b)
1.2.2.2 Effector binding domain

The Helix-turn-helix motifs of FFRPs are connected through a small hinge to its C-terminal domain that has a βαβαβ-fold topology or αβ-sandwich (Brinkman et al., 2000) called as RAM (Regulation of Amino acid Metabolism) domain. It is involved extensively in oligomerization.

**Architecture of the Effector-binding domain** The C terminal domains has βαβαβ-fold topology where the two α-helices are located at one side of the four-stranded antiparallel β-sheet. Two $3_{10}$ helical turns are present, one between β sheet of hinge region connecting N terminal domain and β1 of βαβαβ-fold, and the other between β2 and first α helix of βαβαβ-fold (Leonard et al., 2001).

**Ligand Binding Sites** The C-terminal domain of Lrp-like proteins appears to be involved in ligand-response and activation. Probably, these features are all the result of similar inter and intra molecular rearrangements. Work has been done on the *E. coli* Lrp mutant which either influence DNA binding or leucine response, variation at any of the position L107, D114, M124, L136, Y147, V148 and V149 was found to insensitive to leucine (Platko et al., 1993). Thaw and co-workers (2007) has identified conserved regions by comparative sequence analysis of RAM sequences to two potential ligand binding sites, one between loop connecting strand β3 and β4 of the monomer and the strand β5 of another, together with a second site the mirror image of this at the dimer-dimer interface (Ettema, et al., 2002). Thaw and co-workers (2007) reported the crystal structure of *E.coli* AsnC complexed with asparagine and *Neisseria meningitidis* in complex with leucine and methionine has been reported by Ren, et al in 2007 in both cases ligand binding site was same as proposed by Ettema, et al in 2002 and the entire eight dimer interface contain bound ligand. In crystal structure of DM-FFRP, DM1 forms complex with selenomethionine and isoleucine, ligand-binding site was same as FL-FFRPs but only four of eight dimeric interfaces bound to ligands (Yokoyama, et al., 2007).

Solution studies involving *E. coli* Lrp and a couple of its mutants (Asp113Glu & Leu135Arg), which prevent octamer-hexadecamer transition, revealed the possibility of two kinds of binding sites (Chen et al., 2002). Binding of leucine to a low affinity site induced dissociation of hexadecamers in the unmutated protein to octamers; on the other hand, the octameric form and the mutants retained high affinity...
for leucine. The Asp113Glu mutation maps onto same as proposed by Ettema, et al in 2002 where as Leu135 residue apparently generates possibility for a new site.

**ACT versus RAM domain** RAM domain of FFRPs reveals a $\beta\alpha\beta\alpha\beta$ fold that is strikingly similar to that of the ACT domain, a ubiquitous allosteric regulatory domain of many metabolic enzymes (Ettema, et al, 2002). Despite the structural similarities between the RAM domain and ACT domain the effector-binding sites in these domains seems to be different. On the basis of initial alignment done by Arvind and Koonin (Arvind et al., 1999) a common characteristic is the presence of a nearly invariant glycyl residue at the conserved loop between the first $\alpha$ strand and the first $\beta$ helix that coincided with the binding site for L-serine in phosphoglycerate dehydrogenase (Figure 1.4a). Structural analysis reveals that contact between two ACT domains is mediated via the $\alpha2$ and $\beta3$ interface resulting in an eight standed anti-parallel $\beta$ sheets.

On the other hand proteins containing RAM domain do not have a glycyl residue in this location and this absence was found to be coincide with the presence of another highly conserved glycyl residue in the loop between the $\beta2$ and $\beta3$ strands. The dimer interface is mainly formed by interactions between the anti parallel $\beta$ sheets that are facing each other, forming an antiparallel $\beta$ barrel like structure. Although the RAM domain proteins share many features of the ACT domain alignment, the $\beta2$ - $\beta3$ loop area does appear to be rather distinctive for the RAM grouping. By assuming that the RAM and ACT domains originated from a common ancestor, these observations suggest that their ligand-binding sites have evolved independently. Both domains appear to play analogous roles in controlling key steps in amino acid metabolism at the level of gene expression as well as enzyme activity.
Figure 1.4 (a) ACT domain of 3-phosphoglycerate dehydrogenase (1PSD), with bound serine between α1 and β1. (b) RAM domain of *E. coli* AsnC with bound asparagine, between loop connecting β2 and β3.
### 1.2.3 Types of assemblies

Structurally characterized member of FFRPs are found to exit in different assemblies, which ranges from lower to higher order oligomer, open-closed, ring-helix and heteromeric associations. All type of assemblies are found in different FFRPs. Some common are described here as-

**Dimer** Basic structural unit of all the FFRPs are a dimer (Figure 1.5a). The major interaction involves in the dimeric interface are either hydrophobic interactions or hydrogen bonding and are divided into three groups. First, a hydrophobic core that are held together by interactions between residues in strands β2, β3, β4 and β5 from each monomer and, in addition, the β-sheets are extended to form five-stranded antiparallel β-sheets by main chain hydrogen bonding of strand β6 to strand β3 in the other monomer. Secondly, extensive hydrogen bonding interactions can be seen in the antiparallel β-ribbon formed by the β1 strands from both subunits. The third region of contact is hydrophobic in character and is formed between the N- and C-terminal domains of symmetry-related subunits (Leonard et al, 2001)

**Octamer** Hydrophobic nature of interactions at four dimer-dimer interface leads to octamer formation, with main interaction forms between α5 and strand β5 of one dimer pair, with residues leading into helix α4 and those in a loop between strands β3 and β4 of another dimer (Figure 1.5b, Thaw et al, 2006).

**Closed octameric assembly** Crystal structure of LrpA (1I1G) (Leonard et al, 2001), *E.coli* AsnC (2CG4) and *Bacillus anthracis* LrpC (2CFX) (Thaw et al, 2006), *Neisseria meningitidis* NMB0573 (2P5V) (Ren et al, 2001) and DM1 from *Pyrococcus sp.* OT3 (2Z4P) (Okamura et al., 2007) and TTH0845 from *Thermus thermophilus* (2DJW) (Nakano et al., 2006) exhibit a closed octameric assembly where dimer generate octamer through crystallographic symmetry operations

**Open octameric assembly** *E. coli* Lrp when crystallized in presence of DNA, crystal structure shows an open and linear array of four dimmers, possible explained to be in response of DNA which alters the relative spatial orientation of the DNA binding domain (Figure 1.5c, de Los Rios et al., 2007).

**Helical association** Dimers of FL11 are assembled into continuous cylinders with a right-handed helicity, extended from one end of the crystal to the other for
millimetres. Each monomer forms an asymmetric unit, and there are six dimers in each helical turn of 47.4Å (Figure 1.5d & e).

**Heteromeric association** Solution studies in presence of arginine for DM1 protein, when excessive of FL11 protein added changes the gel elution profile, creating new peak whose molecular weight intermediate of two eluted fraction under normal condition. This molecular weight of heteroassembly explained in terms of association of two DM1 dimers with two FL11 dimers (Okamura et al, 2007).
Figure 1.5 (a) Dimer (2GQQ), (b) Octamer of LrpA (ring), (c) Octamer of *E. coli* Lrp (open), (d) and (e) helix structure of FL11, showing possible direction of DNA helix (d & e adapted from Koike *et al.*, 2003)
1.2.4 Ligand mediated transition of assemblies

Many FFRPs interact with small molecules (ligands or co-regulators), mostly amino acid, which sometimes changes the quaternary association eg- *E. coli* Lrp changes its association state from hexadecamer to octamer in presence of leucine (Chen and Calvo, 2002). In the presence of lysine, FL11 assembles to octamer (Yokoyama *et al*, 2007 & Okamura *et al*, 2007) Arginine facilitates heteroctamerization by two DM1 and two FL11 dimers (Okamura *et al*, 2007).

Solution studies along with the crystal structure of DM1 show that isoleucine stabilizes the octamer formation where as metheonine destabilizes the same assembly. Gel filtration analysis of DM1 leads to an order of amino acids from most assembling to disassembled state as I/V/(L,T,F,A)/R/M/C (Okamura *et al*, 2007 & Sakuma *et al*, 2005).

1.2.5 Proposed mechanisms of Transcription regulation by FFRPs

Possibility of existence of different oligomeric states i.e. an increase in number of DNA binding domains will relax the ideality of binding sites, increasing their deviation. Only a large assembly might recognize two ideal sites positioned far away. Interaction of FFRPs with different type of ligands may either stabilize or destabilize particular type of assembly that may lead to discrimination between a numbers of promoters in response to the environment, mediated by ligand.

Yokoyama and co-workers in 2006 gave a schematic summary (Figure 1.6) for the mechanism of regulation in response to set of ligands signalling types to different environmental changes. According to them ligand interacts with different types of FFRP assemblies and results in their stabilization or destabilization. Ligands leads to stabilizing assemblies or destabilizing assemblies are shown here by shadowed in shapes or in black with thorns preventing the formation of the assemblies. In response to environmental changes; A represent binding of dimer to promoter where number of binding site are two, similarly B represent regulation from FL11 type of helix forming protein where six dimer are present per pitch and C type of regulation is possible for FFRPs form octameric assembly.
Figure 1.6 Figure representing a proposed mechanism of regulation. Drawn figure show that the process proceeds from top left (the environment outside the organism) to bottom right (the metabolism inside the organism).

(Figure adapted from Yokoyama et al. 2006)
E and F represents different type of hetero assembly like FL11 and DMI (Okamura et al., 2007) or even E. coli Lrp and AsnC, FL9 and DMI from Pyrococcus sp OT3 may lead in change of DNA binding specificity. The DM-disc will not bind to any promoter, but its formation will affect other assemblies, e.g. by preventing formation of heterodiscs, thereby indirectly accelerating formation of FL-FFRPs assemblies (Yokoyama et al. 2006).

"Heterotrophic or Autotrophic"/ "Growth or Rest" Regulation

Regulatory mechanisms which are governed by Lrp/AsnC kind of regulators are more or less to be of autotrophic or heterotrophic in case of E. coli Lrp than growth or rest of FL11 of Pyrococcus OT3.

E. coli Lrp is the best characterized in all FFRPs in biological terms. Lrp senses the presence of rich nutrition by interacting with leucine, and functions as a transcriptional repressor as well as an activator and regulates 4070 transcription units. (Calvo et al., 1994; Wagner, 2000 & Newman et al., 1995). E. coli are known to regulate 10% of all genes in E. coli. By blocking the approach of RNA polymerase or the extension of the mRNA by binding of a repressor to the promoter, catabolic pathways functioning in heterotrophic metabolism are repressed and the biosynthetic pathways functioning in autotrophic metabolism are activated by Lrp, where an activator needs to bind upstream of the promoter to interact with RNA polymerase directly or indirectly through another factor. Depending upon the nutritional condition, the number of Lrp dimers present per cell changes. The number of Lrp dimer decreased from 3000 to 2000 when shift to rich media from minimal media, signifies the increase in number with depletion of nutrition, therefore enhancing activation of autotrophic process and repressing heterotrophic pathways.

Growth or rest types of regulation of FFRPs are explained insight of where 7 out of 14 FFRPs interact with amino acids and another FFRPs possibility does in response to amino acid availability. Out of 1025 transcription units coded by Pyrococcus OT3, 271 units are believed to be under control of FL11, which function as a transcriptional repressor (Yokoyama et al., 2007). Sensing rich nutrition as a high concentration of lysine, FL11 forms an octamer and terminate the transcription of fl11 gene and biosynthesis of lysine by binding to the promoter cooperatively. With the decrease in concentration of FL11, transcription of other metabolic which are under control of FL11 are de-repressed, there by activating catabolism of amino acids and

TH-16401
synthesis of ATP, and shifting the metabolism into the feast mode. In the famine mode, sensing the absence of lysine, FL11 disassembles into dimers. So the feast and famine mode of transcription regulation corresponds to growth during feast condition and rest or non growing state under famine conditions (Kawashima et al., 2008).

1.3 Rv3291c:

The Rv3291c gene codes for an Lrp/AsnC (leucine responsive regulatory protein/regulator of asparagine synthase C gene product) type global transcriptional regulator (MtbLrp) Rv3291c has a molecular weight of ~18 kDa and has been found to be up-regulated 15-fold in nutrition starved tuberculosis models designed to mimic the latent/persistent state (Betts et al., 2002). It lies in the genomic region upstream of sigF involved in mycobacterial stress response (Wu et al., 1997) and its expression has also been observed on inhibition of septum formation and is inversely proportional to bacterial growth rate (Slayden et al., 2006 & Landgraf et al., 1996).

More recently, in Mycobacterium fortuitum, the homologous gene was found to be highly conserved and in vivo screening involving a murine infection model of persistence using this pathogen led to the identification of a mutant where this gene was disrupted. It was found that the mutation affected the ability of the pathogen to persist in the kidney in the model (personal communication, Dr R. Srivastava, Central Drug Research Institute, Lucknow, India). The above data suggest that the protein is important for maintenance/adaptation to long-term persistence although it is not essential for growth under normal conditions (Sassetti et al., 2003). Moreover no mammalian homologs of the protein are known. Rv3291c therefore has the potential to be an important therapeutic target for developing strategies to counter persistence of the disease.