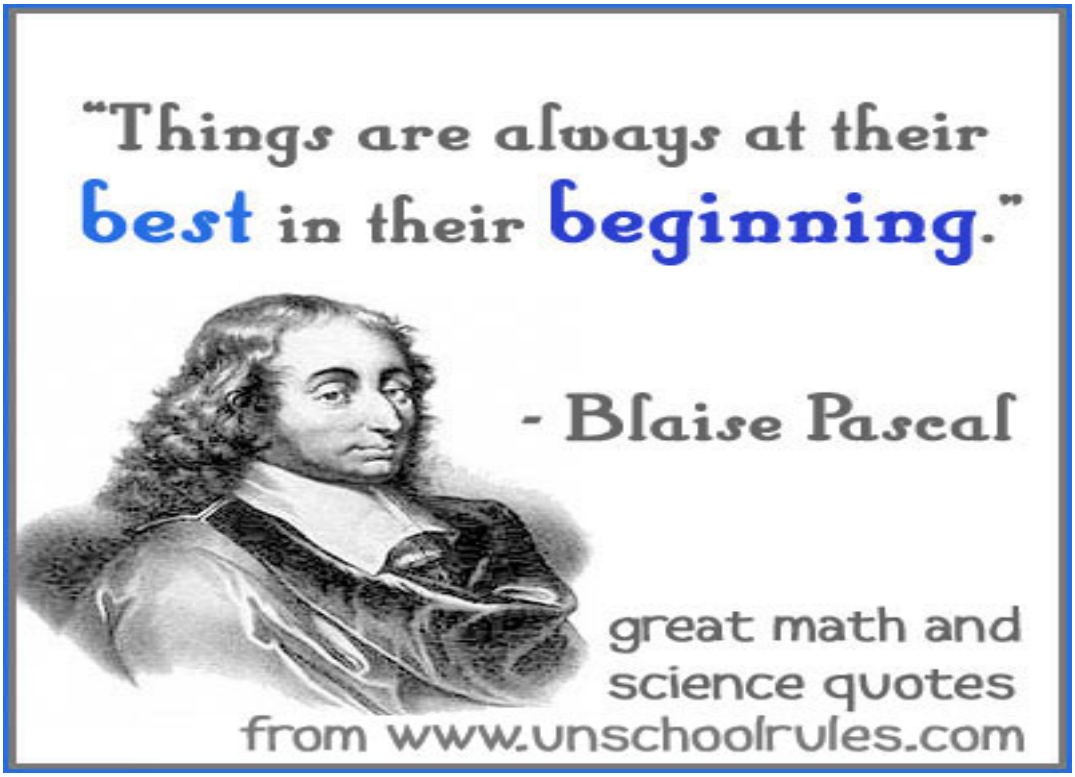

Chapter 1
Introduction and Review of Literature

2.1 Cyanobacteria

Cyanobacteria (formerly called Blue–green algae) are a large group of Gram-negative prokaryotes that perform oxygenic photosynthesis. The initiation of oxygenic photosynthesis in cyanobacteria more than 2.4 billion years ago elicited intense changes in Earth's geochemistry and biosphere (Nickelsen and Rengstl, 2013; Lyons et al., 2014). The most significant event in primeval Earth was escalation in the level of molecular oxygen in the atmosphere **(Fig. 1.1)**. By facilitating energy production via respiration, O_2 provided the basis for the development of complex aerobic life forms and almost all biomass on Earth. Chloroplasts of higher plants are believed to have evolved through endosymbiosis of cyanobacteria and plant cells (Falcon et al., 2010). Cyanobacteria are the only source of nitrogen biofertilizer derived from solar energy and they are the major nitrogen fixers in the oceans of this planet (Capone and Carpenter, 1982). Some cyanobacteria can grow heterotrophically, however, majority of them are characterized by their dependence on photosynthesis, which generates ATP and reducing equivalents in the form of NADPH or reduced ferredoxin that is used in the fixation of CO_2 and the assimilation of other nutrients including nitrogen (Flores and Herrero, 2010).

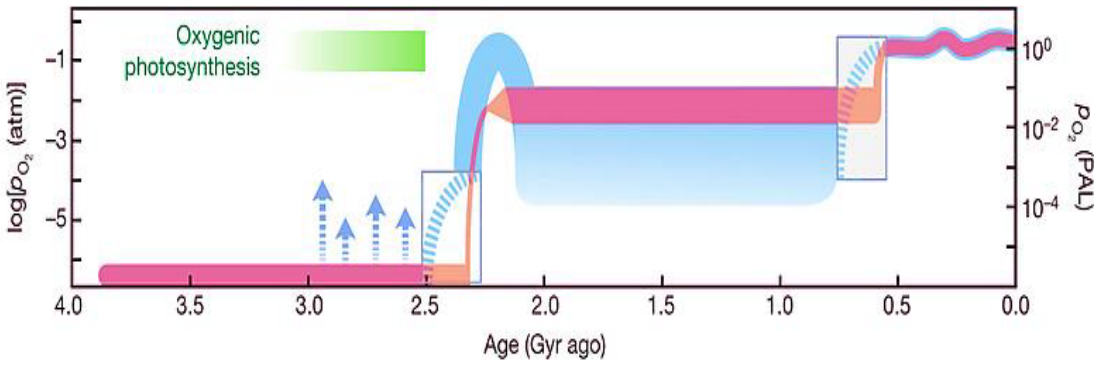

Fig. 1.1 Evolution of Earth's atmospheric oxygen content through time (Lyons et al., 2014).

Cyanobacteria exhibit different morphological features ranging from unicellular forms to filamentous. Many of them can fix atmospheric nitrogen. Nitrogen metabolism of cyanobacteria, especially nitrogen fixation, has been significant since its regulation is coordinated with photosynthesis. However, photosynthesis and nitrogen fixation are incompatible processes because nitrogenase is inactivated in presence of oxygen liberated during the photosynthesis. To avoid this problem, unicellular cyanobacterium *Cyanothece* sp. strain ATCC 51142 **(Fig. 1.2a)**

fixes CO_2 during the day and nitrogen at night (Toepel et al., 2008; Welsh et al., 2008), whereas nitrogen fixation in some of the filamentous forms, such as *Anabaena sp.* PCC 7120 **(Fig. 1.2b)**, occurs in specialized cells called heterocysts **(Fig. 2c)** (Kumar et al., 2010). The heterocysts provide the anaerobic environment for nitrogen fixation.

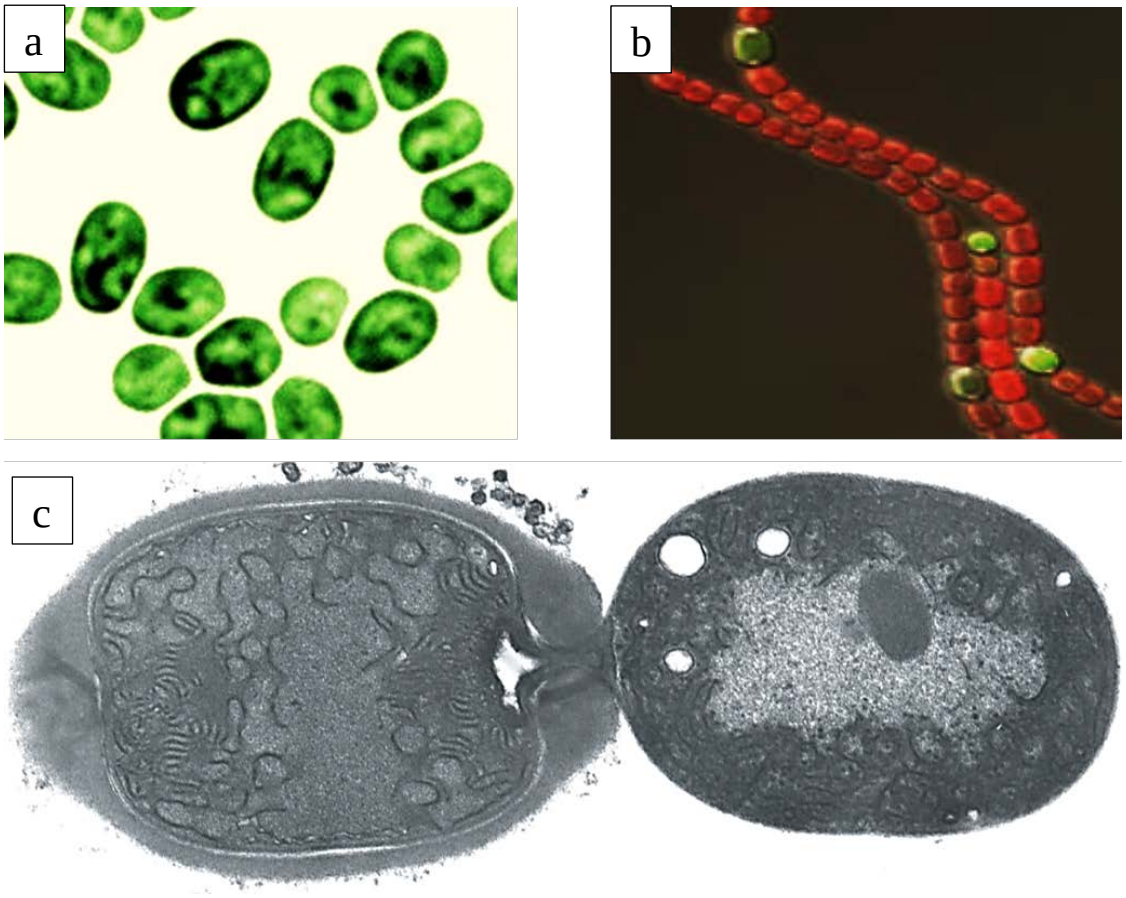

Fig. 1.2 Different morphological forms in cyanobacteria. (a) **Unicellular *Cyanothece sp.* 51142.** (b) **Filamentous *Anabaena sp.* PCC 7120.** (c) ***Anabaena sp.* PCC 7120 heterocyst (left) and vegetative cell (right) (Flores and Herrero, 2010; Kumar et al., 2010; Sanderson, 2010).**

2.1.1 *Anabaena* PCC 7120

Anabaena sp. strain PCC 7120 is a filamentous, heterocyst-forming obligate autotroph cyanobacterium capable of performing oxygenic photosynthesis (Kaneko et al., 2001; Flores and Herrero, 2010). This filamentous cyanobacterium protect extremely oxygen-sensitive nitrogenase by spatially splitting nitrogen-fixation and oxygenic photosynthesis in two different cell types, the oxygen evolving vegetative cells and the nitrogen-fixing heterocyst. Deprivation of combined nitrogen initiates an intricate signal transduction cascade resulting in morphological, biochemical and

genetic changes in *Anabaena* PCC 7120. This triggers differentiation of about 5-10 % of vegetative cells in to heterocyst at semi-regular intervals along the filament.

2.1.1.1 *Anabaena* PCC 7120 Heterocyst formation

Heterocysts are characteristically distinguishable from vegetative cells in having larger and rounder shape, reduced pigmentation, thicker cell envelopes, and prominent cyanophycin granules at poles adjacent to vegetative cells **(Fig. 1.3)** (Kumar et al., 2010). Heterocyst provides a suitable microaerobic environment, which is tolerated by nitrogenase (Haselkorn et al., 1978; Flores and Herrero, 2010; Kumar et al., 2010). They maintain a very low internal partial oxygen pressure (pO_2) various means (Yoon and Golden, 1998). Oxygen producing PSII is inactivated in heterocysts but remains active other vegetative cells (Wolk, 1996). Additionally, molecular oxygen is reduced to water by respiration through heterocyst specific terminal respiratory oxidases (Valladares et al., 2003; Walsby, 2007). In the absence of PSII and ribulose 1,5-bisphosphste carboxylase oxygenase, heterocysts are unable to photosynthetically fix CO_2 and hence depend upon neighboring vegetative cells for carbon source presumably in the form of sucrose (Wolk, 1968; Curatti et al., 2002; Marcozzi et al., 2009). Heterocysts are devoid of glutamine-2-oxoglutarate amidotransferase but possess high levels of glutamine synthase. Vegetative cells provide glutamate to the heterocysts where glutamate gets converted to glutamine and other amino acids.

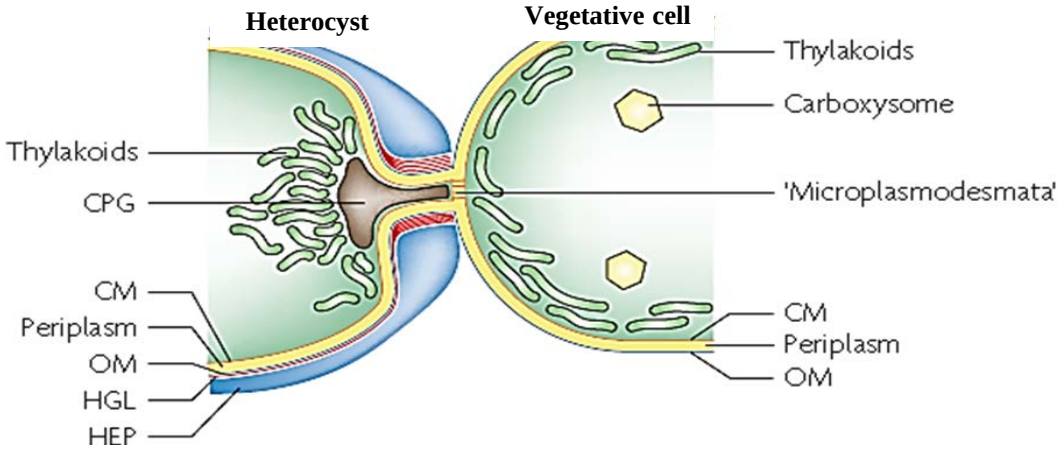

Fig. 1.3 Morphological changes during heterocyst formation (Flores and Herrero, 2010).

In return, heterocysts supply fixed nitrogen to vegetative cells (Meeks et al., 1988; Kumar et al., 2010). Morphological changes in heterocyst include the formation of

two additional envelope layers: an inner "laminated" layer composed heterocyst-specific glycolipids (HGL) and an outer polysaccharide layer (HEP). The polysaccharide layer consists of a well-defined homogeneous inner layer and an external fibrous layer (Wolk et al., 1988; Nicolaisen et al., 2009a; Nicolaisen et al., 2009b). HGL envelope provides a barrier to the entry of oxygen and external layer of HEP protects HGP layer from physical damage.

During nitrogen step down, elevation of intracellular 2-oxoglutarate (2-OG) initiates heterocyst differentiation. Complete formation of heterocyst takes about 24 h from the nitrogen step down where approximately 10 % of chromosomal genes are regulated **(Table 1.1 & 1.2)** (Ehira, 2013). Nitrogen stress response is well orchestrated by NtcA, a global cyanobacterial transcriptional regulator, whose activity is known to be positively modulated by 2-OG (Herrero et al., 2001; Flores and Herrero, 2010; Zhao et al., 2010). NtcA is known to bind GTA-N_8-TAC consensus sequence (Chastain et al., 1990). NtcA mutants of *Anabaena* PCC 7120 displayed pleiotropic effects and was found to be essential transcriptional factor in heterocyst differentiation (Wei et al., 1994).

2.1.1.2 Nitrogenase

Nitrogen-fixers catalyze the reduction of dinitrogen (N_2) to two molecules of ammonia (NH_3) via nitrogenase enzyme (**Fig. 1.4)**, the major contribution of fixed nitrogen to the biogeochemical nitrogen cycle (Seefeldt et al., 2009). Reduction of N_2 by this enzyme involves transient interaction of Fe protein and MoFe.

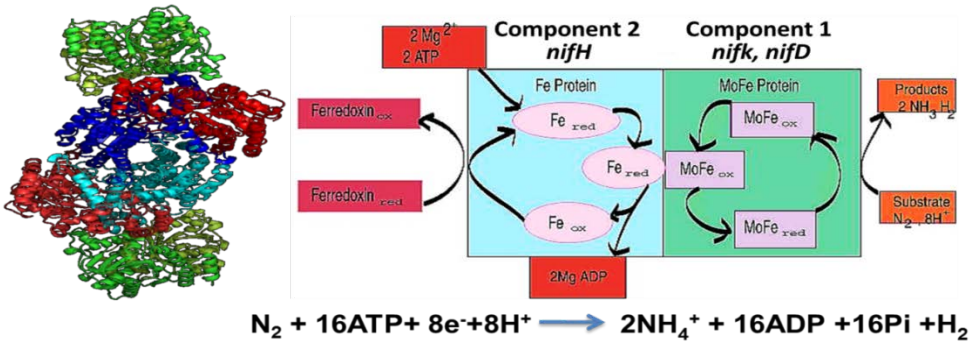

Fig. 1.4 Structure and mechanism of nitrogenase showing reduction of N_2 to two molecules of NH_3.

Table 1.1 *Anabaena* PCC 7120 genes upregulated during heterocyst formation (Ehira, 2013).

ORF	Gene name	Product
all0177		flavoprotein
all0178		flavoprotein
all0217		similar to pathogenesis related protein
all0218	*pdxJ*	pyridoxal phosphate biosynthetic protein
alr0348	*ndhD*	NAD H dehydrogenase subunit 4
asr0485	*pipX*	PII interaction protein X
all0571	*cphB2*	cyanophycinase
all0687	*hupL*	[NiFe] uptake hydrogenase large subunit
all0688	*hupS*	[NiFe] uptake hydrogenase small subunit
alr0692		similar to NifU protein
alr0694	*hypF*	hydrogenase maturation protein
asr0695	*hupC*	hydrogenase expression/formation protein
alr0696	*hupD*	hydrogenase expression/formation protein
asr0697		probable 4-oxalocrotonate tautomerase
alr0698	*hupE*	hydrogenase expression/formation protein
alr0699	*hupA*	hydrogenase expression/formation protein
alr0700	*hupB*	hydrogenase expression/formation protein
all0769		acetyl-CoA synthetase
all0808		ABC transporter, DevC homolog
all0809		ABC transporter, DevB homolog
all0827		putative oxidoreductase
alr0870	*ndhD4*	NAD H dehydrogenase subunit 4
all0948		heme O synthase
alr0992		ammonium transporter
alr0996		protease
alr1050	*pgi*	glucose-6-phosphate isomerase
all1126		NAD H dehydrogenase
all1127		NAD H dehydrogenase
alr1336	*pknH*	serine/threonine kinase
alr1337		probable ribosylglycoyhdrolase
alr1404		serine acetyltransferase
alr1407	*nifV1*	homocitrate synthase
asr1409	*nifT*	NifT protein
all1420		probable short chain dehydrogenase
all1430	*fdxH*	heterocyst-specific [2Fe-2S] ferredoxin
all1431	*hesB*	HesB protein
all1432	*hesA*	HesA protein
all1433	*nifW*	nitrogen fixation protein
all1436	*nifX*	nitrogen fixation protein
all1437	*nifN*	nitrogenase FeMo cofactor biosynthesis scaffold protein
all1438	*nifE*	nitrogenase FeMo cofactor biosynthesis scaffold protein
all1440	*nifK*	nitrogenase MoFe protein subunit beta
all1454	*nifD*	nitrogenase MoFe protein subunit alpha
all1455	*nifH*	nitrogenase iron protein
all1456	*nifU*	nitrogen fixation protein
all1457	*nifS*	nitrogenase cofactor synthesis protein
all1488	*dnaJ*	chaperone protein
all1516	*fdxN*	ferrodoxin-like protein
all1517	*nifB*	nitrogen fixation protein
alr1521	*invA*	alkaline invertase
all1553		NAD H dehydrogenase
alr1633		sulfate permease
alr1635		sulfate permease
all1880		an acsF-like protein
alr1963	*clpB*	endopeptidase Clp ATP-binding chain
alr2046		putative potassium/proton antiporter
all2289		glucosyltransferase
alr2493	*ycf16*	ABC transporter ATP-binding protein
all2512	*patB*	transcriptional regulator
asr2513	*fdxB*	ferredoxin

ORF	Gene name	Product
alr2514	*coxB2*	cytochrome *c* oxidase subunit II
alr2515	*coxA2*	cytochrome *c* oxidase subunit I
alr2516	*coxC2*	cytochrome *c* oxidase subunit III
all2521		cysteine synthase
alr2526		luciferase-alpha subunit
asl2551		transcriptional regulator
all2563	*talB*	transaldolase
all2564	*pyk1*	pyruvate kinase
all2566	*gap1*	glyceraldehyde-3-phosphate dehydrogenase
alr2731	*coxB3*	ARTO, homolog of cytochrome *c* oxidase subunit II
alr2732	*coxA3*	ARTO, homolog of cytochrome *c* oxidase subunit I
alr2734	*coxC3*	ARTO, homolog of cytochrome *c* oxidase subunit III
alr2803	*nifJ1*	pyruvate : flavodoxin oxidoreductase
all2910		carbonic anhydrase
alr3125		heme oxygenase
alr3126		oxygen independent coprophorphyrinogen III oxidase
alr3287		nitrate transport protein
alr3356		similar to phytochrome
alr3358		3-hydroxyacid dehydrogenase
alr3701		defective type II restriction site-specific deoxyribonuclease
alr3710	*devB*	heterocyst specific ABC-transporter, membrane fusion protein
alr3712	*devA*	heterocyst specific ABC-transporter, ATP-binding subunit
all3765		two-component hybrid sensor and regulator
alr3956	*ndhF1*	NAD H dehydrogenase subunit 5
alr3997		serine/threonine kinase
all4008	*pyk2*	pyruvate kinase
all4018	*opcA*	putative OxPPCycle protein
all4019	*zwf*	glucose 6-phosphate dehydrogenase
all4021	*fbp*	fructose 1,6-bisphosphatase
alr4029		similar to vitamin B12 transport protein
all4052	*tktB*	transketolase
all4121	*petH*	ferredoxin-NAD P(+) reductase
alr4249	*sigE*	RNA polymerase group 2 sigma subunit
all4312	*nrrA*	nitrogen-responsive response regulator
alr4619		similar to pathogenesis related protein
alr4849		methanol dehydrogenase regulatory protein
all4906		phosphoglycerate mutase
alr5178		molybdate-binding periplasmic protein
alr5275	*gnd*	6-phosphogluconate dehydrogenase
all5341	*hglT*	Heterocyst specific glycolipid synthase
all5345		oxidoreductase
all5346		ABC transporter, DevC homolog
all5347		ABC transporter, DevB homolog
alr5351	*hglE*	heterocyst glycolipid synthase
alr5354	*hglD*	heterocyst glycolipid synthase
alr5355	*hglC*	heterocyst glycolipid synthase
alr5357	*hetM*	polyketide synthase
alr5358	*hetN*	ketoacyl reductase
alr5362		sugar ABC transporter ATP binding protein

Table 1.2 *Anabaena* PCC 7120 genes down regulated during heterocyst formation (Ehira, 2013).

Gene		Function
asl0137	*psbTc*	photosystem II protein
alr0523	*pecB*	phycoerythrocyanin beta chain
all0646		hypothetical protein
asl0846	*psbH*	photosystem II protein
asl0883	*psbM*	photosystem II protein
alr1524	*rbcL*	RuBisCO large subunit
alr1525	*rbcX*	possible RuBisCO chaperonin
alr1537		hypothetical protein
alr3421	*petB*	plastoquinol-plastocyanin reductase, cytochrome b_6
alr3608		similar to endoglucanase
alr4123	*prk*	phosphoribulokinase
alr4290	*psbD*	photosystem II reaction center D2 protein
alr4291	*psbC*	photosystem II CP43 protein
all4499		hypothetical protein
alr4866	*psbAI*	photosystem II reaction center D1 protein

protein**.** The process consumes 16 ATP molecules, 8 protons and 8 electrons to produce two molecules of NH_3.

The MoFe protein is a $\alpha_2\beta_2$ tetramer and the Fe protein is a γ_2 dimer. A MoFe protein binds two Fe proteins, with each αβ-unit functioning as a complete catalytic unit. In *Anabaena* PCC 7120, the α subunit of MoFe protein is encoded by *nifK*, whereas *nifD* codes for β subunit. In case of Fe protein, γ subunit is encoded by *nifH.*

Oxygen interferers with the nitrogenase catalysed nitrogen fixation at different levels (Goldberg et al., 1987). Genetically, oxygen represses the genes involved in nitrogensase synthesis. Structurally, oxygen damages nitrogenase and reversibly inhibits the enzyme by interfering in the redox chemistry of the protein. *Anabaena* PCC 7120 however addresses this problem by heterocyst specific expression of genes involved in nitrogenase synthesis.

2.1.1.3 Genomic Rearrangement in Anabaena PCC 7120

Complete genome of *Anabaena* PCC 7120 consists of a single chromosome (6,413,771 bp) and six plasmids, designated as pCC7120*α* (408,101 bp), pCC7120*β* (186,614 bp), pCC7120*γ* (101,965 bp), pCC7120*δ* (55,414 bp), pCC7120*ε* (40,340 bp), and pCC7120*ζ* (5,584 bp) (Kaneko et al., 2001). *Anabaena* genome is interrupted by three DNA elements namely, *nifD*, *hupL* and *fdxN* element depending on the gene which they interrupt. These elements vary in size as *nifD* element is of 11.2 kb, *fdxN* element is of 59 kb and *hupL* element is of 9.4 kb **(Fig. 1.5 & Table 1.3)**. The *nifHDK* and *nifBfdxN nifSU* operons are located next to each other while the *hupL* operon is located >700kb away on the *Anabaena* chromosome (Kuritz et al., 1993). The interrupting elements get excised out precisely during heterocyst differentiation by means of site-specific recombination events, resulting in gene rearrangement and formation of functional operons.

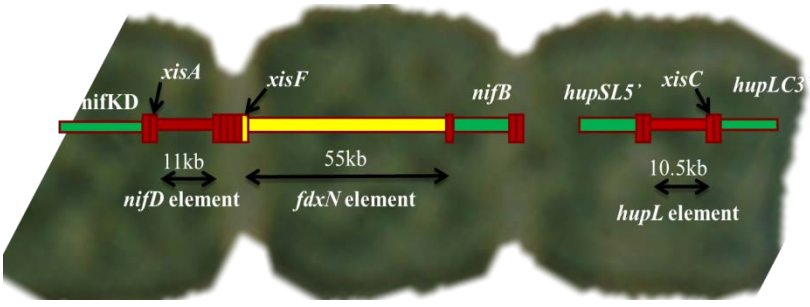

Fig. 1.5 Arrangement of DNA elements in vegetative cells of *Anabaena* PCC 7120.

The excised circles persist in the heterocysts with no known function (Haselkorn, 1992; Apte, 1994a).

The first genetic element to be sequenced in *Anabaena* PCC 7120 was 11.2 kb *nifD* element interrupting the *nifD* gene (Lammers et al., 1986), which codes for α subunit of dinitrogenase, a part of nitrogenase enzyme complex responsible for nitrogen fixation. The *nifD* element contains *xisA* gene proximal to *nifK* gene, whose product was responsible for its site-specific excision. It also contains several other uncharacterized ORFs and cyp110, which codes for cytochrome 450 like protein. Many other cyanobacteria also contain the *nifD* element however the size varies greatly from 4 – 24 kb (Henson et al., 2011).

Anabaena variabilis contains a 11 kb element which shows very high similarity with that of PCC 7120 but the 24 kb *nifD* element from *Nostoc puctiforme* is quite different.

Rearranged element	***nifD* element**	***fdxN* element**	***hupL* element**
Size (bp)	11,278	59,428	9,419
Interrupted gene	*nifD*	*fdxN*	*hupL*
Size of Direct repeats (bp)	11	5	16
Target Sequence	GGAGTAATCCG	TATTC	CACAGCAGTT ATAATG

Table 1.3 Disrupting DNA elements of *Anabaena* PCC 7120 and their characteristics.

Excision of *nifD* element is carried out by XisA, coded by *xisA* gene, present at 3' end of *nifD* element by site specific recombination involving 11bp target site flanking the element (Lammers et al., 1990; Karunakaran et al., 2008). This excision is required for the nitrogen fixation. XisA is necessary for the excision. The predicted ORF of the *nifD* element encodes a polypeptide for 480 amino acids. After excision a new ORF is created by the fusion of the *nifD* ORF to *nifK* 5' flanking sequence. This new ORF codes for 497aa, of which C-terminal 43 residues replaced 26 C-residues of vegetative cell ORF. The new ORF shows similarity to *Rhizobium* NGR 234 of *Parasponia* α subunit of nitrogenase (Golden *et al*., 1985).

***hupL* element:** 9.4 kb *hupL* element disrupts the *hupL* gene, which codes for membrane bound uptake hydrogenase (Carrasco et al., 2005; Henson et al., 2011). It undergoes rearrangement simultaneously with other rearrangements. DNA hybridization studies suggested that three strains of *Anabaena* possess *hupL* element. Main function of the uptake hydrogenase is in the creation/maintenance of micro-oxic condition. However, mutation or absence of this enzyme does not prevent the nitrogen fixation, as in case of *fdxN* element and *nifD* element, but could decrease its efficiency.

***fdxN* element:** 55.2 kb *fdxN* element resides in *fdxN* gene, part of *nifB-fdxN-nifS-nifU* operon (Mulligan and Haselkorn, 1989; Ramaswamy et al., 1997). The presence of 11 and 55 kb elements in *Nostoc* Mac, *Anabaena cylindrica* and *Anabaena* M131 are also reported (Carrasco and Golden, 1995). The Marine cyanobacterium *Anabaena* CA lacks both the elements. Mutation in this element results in the loss of nitrogen fixation as the operon could not be functional. *xisF* gene is responsible for the excision of the *fdxN* element, however it alone is not capable of doing recombination, it requires two other gene products i.e. XisH and XisI. XisF is known to be sufficient for excision while XisH and XisI appears act as accessory proteins similar to integrase lambda.

Three site specific recombinases, XisA, XisC and XisF, are responsible for the excision events in *Anabaena* PCC 7120 (**Fig. 1.6)**. They recognize a specific sequence flanking the harboring element **(Table. 1.4)** - XisA recognizes CGGAGTAATCC 11base pair direct repeat sequence, XisF recognizes TATTC 5 bp sequence and XisC recognizes CACAGCAGTTATATGG 16 base pair sequence (Ramaswamy et al., 1997; Carrasco et al., 2005; Henson et al., 2008). All three recombinases are specific and are necessary for excision as mutation in them results in the loss of excision. They are not involved in heterocyst formation as deletion mutants of these excisases form normal heterocysts (Brusca et al., 1990; Kumar et al., 2010). The deletion of the 55 kb element results in the formation of the *fdxN* ORF and presumably allows for the expression of the downstream genes in the *nifB-fdxN- nifS-nifU* operon by the *nifB* promoter (Mulligan et al., 1988; Mulligan and Haselkorn, 1989). Excision of *fdxN* occurs independently form the 11kb *nifD* element excision (Golden et al., 1988; Golden and Wiest, 1988; Carrasco et al., 1995; Golden and Yoon, 2003).

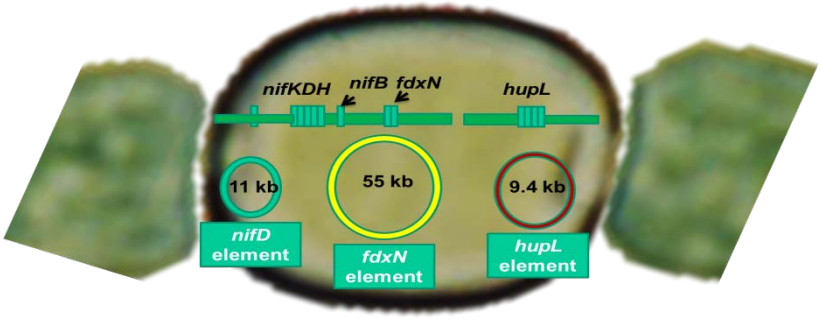

Fig. 1.6 Excision of DNA elements in *Anabaena* PCC 7120 heterocyst.

Recombinase	Excisase A	Excisase F	Excisase C
Gene	*xis A*	*xis F*	*xis C*
ORF	1419 bp	1545 bp	1497 bp
Amino Acids	472	514	498
Estimated Mass (kDa)	52	57	55
Homology	**Integrase**	Resolvase	**Integrase**

Table 1.4 Characteristics features of *Anabaena* PCC 7120 site specific recombinases.

The HupL polypeptide is homologous to the large subunit of [NiFe] uptake hydrogenase. *hupL* is expressed in a manner similar to other nitrogen-fixation genes. *hupL* messenger RNA was detected only during late stages of heterocyst differentiation. An ORF, named *xisC* was identified near one end of the *hupL* DNA element is presumed to encode element's site-specific recombinase.

A 1.5 kb ORF of XisC was identified 115 bp inside the right border of the *hupL* element. Comparison with Genbank sequences identified the homology with only XisA protein. XisA protein sequence that was reported earlier (Lammers et al., 1986) has been corrected (Lammers et al., 1990)**.** Although XisC is 25 % larger than the XisA, the amino acid sequence can be aligned along its full length with XisC. The sequence is 65 % similar and 43 % identical and they appear to belong to Intergrase family **(Fig. 1.7)** (Nunes-Duby et al., 1998)**.** The transcription of *xisA/xisC* genes has been difficult to study because of their low levels of expression and transcription start site has not been identified so far for both genes. However, two *Anabaena* PCC 7120 DNA binding proteins NtcA and factor2 have been shown to bind the sequence upstream of *xisA* gene. The NtcA binding sites are not present in the upstream to *xisC.*

Anabaena PCC 7120 uptake hydrogenase is required to improve the efficiency of nitrogen fixation. *Anabaena* PCC 7120 *xisF* recombinase, which is required for excision of *fdxN* element, belongs to the resolvase family and is homologous to *Bacillus subtilis* SpoIVCA site-specific recombinase (Sato et al., 1996). The *hupL* may differ from the *nifD* and *fdxN* rearrangement in that failure to produce hydrogenase would not be expected to block nitrogen fixation in heterocyst nor growth on nitrogen free medium but could only decrease the efficiency of nitrogen fixation (Carrasco et al., 1995).

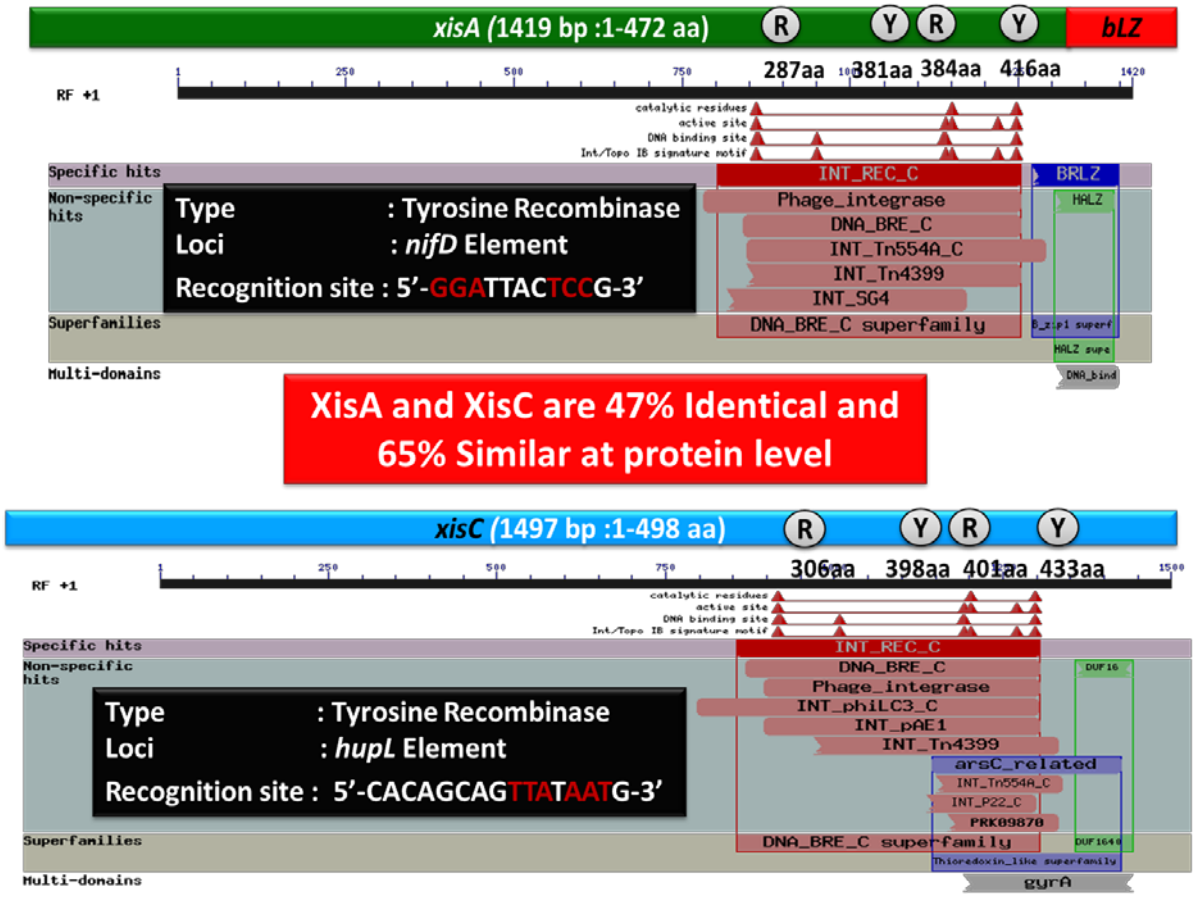

Fig. 1.7 Identity and similarity between XisA and XisC excisases of *Anabanena* PCC 7120 (NCBI).

1.1.1.3.1 *nifD* element rearrangement studies in *E. coli*

XisA mediated excision of *nifD* element has been well-studied in heterologous host *E. coli* wherein rearrangement occurs under nitrogen sufficient conditions at a frequency of about 0.3-11 % depending on the growth medium (Lammers et al., 1986; Karunakaran et al., 2008). *nifD* element excision is enhanced by about 5 fold in M9

minimal medium compared to Luria Bertani (LB) broth in *E. coli* when the strain is wild-type for *recA* (Karunakaran et al., 2008). As opposed to *Anabaena* PCC 7120, *nifD* element excision in *E. coli* DH5α was unaffected by iron-deficiency. Accessory factors, such as RecA and physiological state of *E. coli* play an important role in the excision of *nifD* element. Previous lab study with *nifD* rearrangement in *E. coli* helped in understanding the mechanism of excision and endonuclease activities of XisA protein. P_{xisA}::*gfp* promoter fusion revealed that P_{xisA} activity was increased up to 75 % in *E. coli* strain JM101 in presence *Klebsiella pneumoniae* NifA, which is a transcriptional activator of *nif* genes. Reduction of excision frequency (0.3 %) in *E. coli ntrA*$^{-}$ mutants demonstrated that NifA is involved in enhancing *xisA* gene expression. NtcA, a transcriptional regulator of nitrogen metabolism in *Anabaena* PCC 7120 enhanced the excision of *nifD* element up to 30 % in *E. coli* JM101. These studies indicate the role of transcriptional activators of nitrogen metabolism in controlling *xisA* expression in *Anabaena* PCC 7120. NtcA protein, a global nitrogen stress transcription factor is known to respond to depletion of combined nitrogen sources by allosteric regulation under enhanced intracellular 2-OG concentration which modulates NtcA dependent gene expression (Zhao et al., 2010). The down-regulation of *ccbP* gene in heterocysts, critical to the regulation of Ca^{+2}, depends on functional NtcA. *E. coli* host coexpressing *ntcA* and a *ccbP* promoter-controlled *gfp* genes showed diminished production of GFP, and NtcA activity was found to be modulated by elevating 2-OG levels in culture medium mimicking heterocyst condition (Shi et al., 2006). Hence, *xisA* promoter activity was low, which is expected from lack of NtcA binding due to low 2-OG levels present in combined nitrogen sources. On the other hand, enhancement in *nifD* excision frequency was seen which may be attributed to DNA bending activity of NtcA.

K. pneumoniae NifA transcription factor activates expression of *nif* genes under nitrogen stress. NifL, a signal transduction protein negatively modulates the activity of NifA protein by direct protein-protein interaction in presence of combined nitrogen sources (Filser et al., 1983; Cannon and Buck, 1992; Narberhaus et al., 1995; Grabbe et al., 2001). However, in heterologous host like *E. coli* where NifL protein absent, NifA activity remains independent of concentration of combined nitrogen source (Schmitz et al., 2002). *K. pneumoniae* NifA and *Anabaena* PCC 7120 NtcA proteins are known to activate genes transcribed by σ^{54} and σ^{70} dependent RNA

polymerases, respectively; however both transcription factors have overlapping binding site (Apte, 1994b). Sequence analysis of the promoter region of the *xisA* gene showed the presence of a σ^{54} consensus promoter sequence $GGN_{10}GC$ (P3 promoter, **Fig. 1.8**) in addition to σ^{70} dependent promoters P1 and P2 reported by Lammers *et al.* (1986; 1990). Promoters P2 and P3 are preceded by three tandem NifA/NtcA binding sites. Present study showed that only NifA elevated P_{xisA} activity particularly in stationary phase in *E. coli* while NtcA had no significant effect under similar conditions.

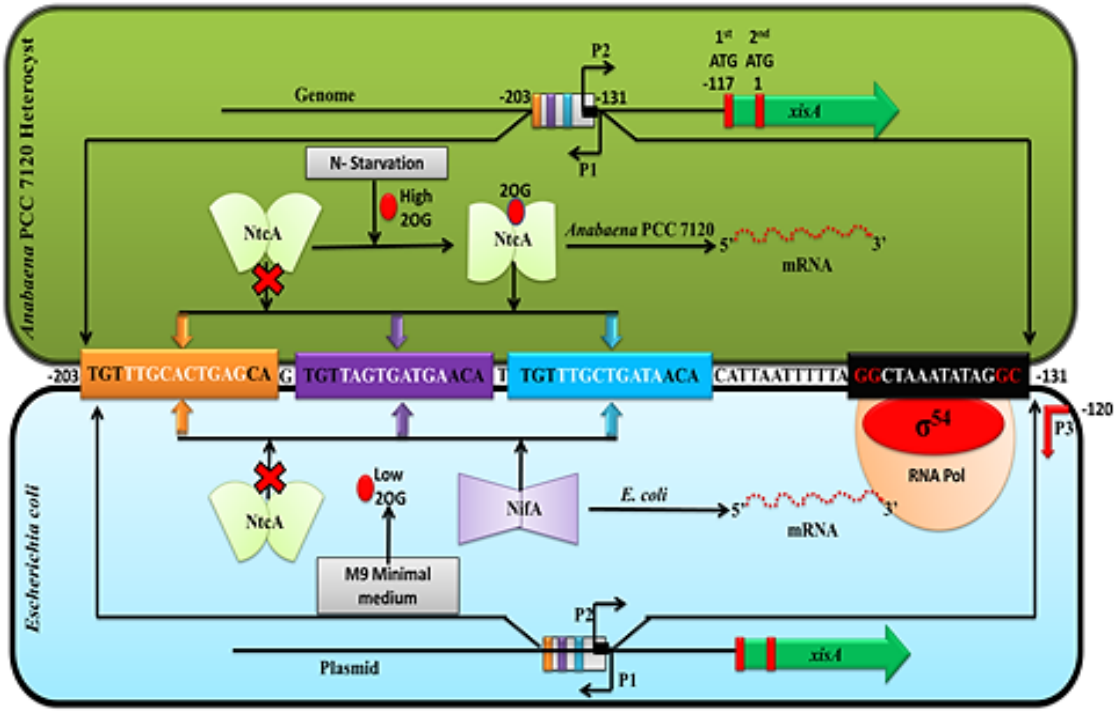

Fig. 1.8 Proposed model for the expression of *Anabaena* PCC 7120 *xisA* gene in *E. coli*. In M9 minimal medium, *xisA* expression occurs from P2 promoter (weak) but shifts to σ54 dependent P3 promoter (strong) in presence of NifA protein. However, NtcA dependent activation of *xisA* gene expression does not occur in *E. coli* from P2 promoter as 2-OG levels are low.

This could be attributed to the constitutive nature of NifA activity and inactivity of NtcA due to low 2-OG levels prevalent under N-sufficient conditions, 2-OG levels are low (Buchanan-Wollaston et al., 1981; Tanigawa et al., 2002).

2.2 Characteristics of enzymes involved in DNA cleavage and recombination.

2.2.1 Site specific recombinases

Site specific recombination (SSR) involves binding, cutting, strand exchange and ligation, followed by release to complete the process (Grindley et al., 2006). The minimum requirement for the SSR involves recombination site and recombinase for

e.g. FLP and Cre recombinase, while other recombinases require additional helper proteins and DNA sequences other than the cross over site. Site-specific recombinase are divided into main two groups depending on the active site amino acid.

1. **Tyrosine recombinase also known as integrase family** (Table 1.5).
2. **Serine recombinase also known as resolvase family** (Table 1.6).

2.2.1.1 Tyrosine Recombinases

Members of the tyrosine recombinase (YR) family of catalyze a variety of sequence-specific DNA rearrangements across all the domains of life, which include the integration and excision of phage genomes into and out of their bacterial hosts (Czurda et al., 2010).

Tyrosine Recombinase	
Recombinase	**Biological function**
λ Int and many other phage integrases	Integration and excision of phage genomes
Int of Tn916/Tn1545	Integration and excision: transposition of circular transposons
IntI	Integration and excision of gene cassettes in integrons
Cre	Excision: dimer reduction in phage P1 plasmids
XerC/D	Excision: dimer reduction in the *E. coli* chromosome as well as in many other bacterial chromosomes and some plasmids
TnpI of Tn4430	Excision: resolution of cointegrates resulting from transposition of Tn4430
FimB, FimE	Inversion: alternation of gene expression (fimbrial phase variation in *E. coli*)
Rci of R64	Inversion of shufflon segments in plasmid R64, producing various forms of pili
XisA, XisC	Excision: for developmentally regulated gene activation in *Anabaena*
Flp	Inversion: for amplification of yeast 2-μm plasmid

Table 1.5 Members of tyrosine recombinase family and their biological functions.

Serine Recombinase	
TnpR of Tn3/γδ and related transposons	Excision: resolution of cointegrates resulting from transposition
Sin of *Staphylococcus aureus*	Excision: dimer reduction in staphylococcal plasmids
ParA of RP4	Excision: dimer reduction in plasmid RP4
Hin	Inversion: alternation of gene expression (flagellar phase variation) in *Salmonella*
Gin, Cin	Inversion: alternation of gene expression (tail fiber proteins) in phages Mu and P1
OrfA of IS607/IS1535	Integration and excision: transposition of the *Helicobacter pylori* element IS607 (and others?)
Int of ϕC31/Bbv1/ϕRv1[a]	Integration and excision of *Streptomyces* and mycobacterial phages
TnpX of Tn4451[a]	Integration and excision: transposition of Tn4451 in *Clostridium*
SpoIVCA (CisA)[a]	Excision: for developmentally regulated gene activation in *Bacillus subtilis*
XisF[a]	Excision: for developmentally regulated gene activation in *Anabaena*

Table 1.6 Members of serine recombinase family and their biological functions.

The phage P1 Cre recombinase **(Fig. 1.9a)** and lambda phage λ integrase **(Fig. 1.9b)** are well-studied YR systems. The YR enzymes bind to two distinct, specific sequences (e.g. attP and attB for λ integrase; two loxP sites for Cre) and catalyze a stepwise exchange of strands between the two sites to generate recombinant products (Grindley et al., 2006). The tyrosine recombinase mediated site-specific recombination reaction begins with cleavage of the recombining sites by conserved tyrosine residues to form covalent 5' phospho-tyrosine linkages to the DNA substrates and release free 3'hydroxyl groups **(Fig. 1.10)**. Alternatively, the strands bearing free 5' hydroxyl groups can be exchanged between recombining sites and subsequent ligation results in formation of a four-way Holliday junction (HJ) intermediate. The HJ intermediate can then serve as the substrate for a second round of cleavage, strand exchange and ligation steps. If the same strands are processed to resolve the junction, the starting sites are re-generated. If the HJ intermediate isomerizes to allow processing of the opposite pair of DNA strands, recombinant products are formed.

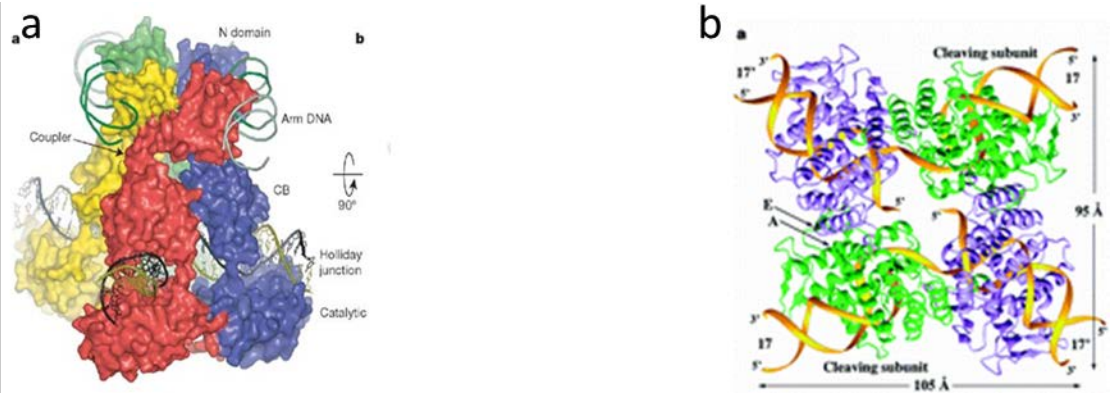

Fig. 1.9 Crystal structure of site specific recombinases. (a) **Cre recombinase (Guo et al., 1997).** (b) **Lambda integrase (Biswas et al., 2005).**

YRs are characterized by conserved catalytic tetrad **R-$(X)_{70-139}$-H/Y-XX-R-$(X)_{29-38}$-Y (Fig. 1.11)** (Esposito and Scocca, 1997). PSI BLAST hits from NCBI RafSeq suggests that primary arginine of active site is conserved for about 91.2 % members while only 1.1 % members have lysine at this position. The second catalytic position is occupied by histidine in about 87.2 % members while in all the cynobacterial tyrosine recombinases including XisA and XisC, tyrosine occupies this position (Esposito and Scocca, 1997). Secondary arginine is highly conserved and only few members of tyrosine recombinases (0.5 %) have a lysine at this position.

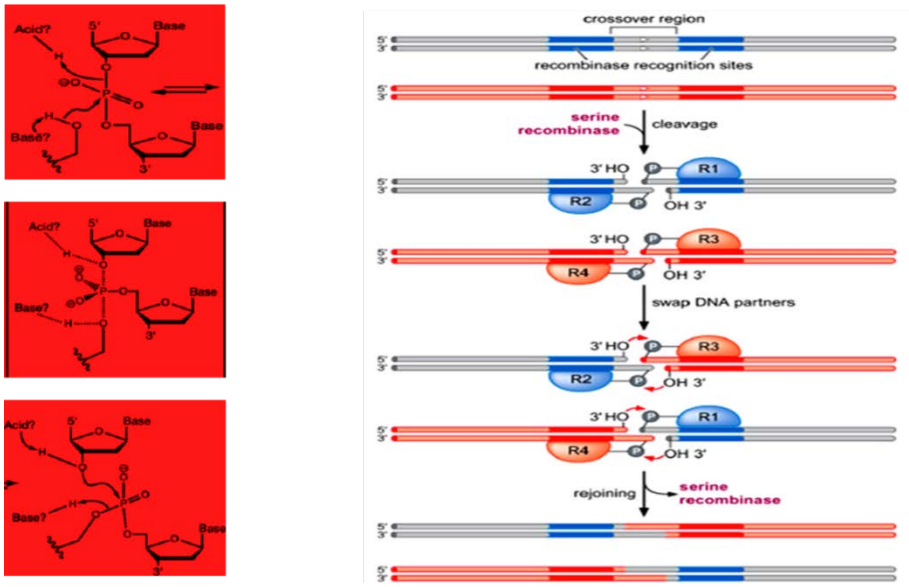

Fig. 1.10 Catalytic Mechanism of tyrosine recombinases (Grindley et al., 2006).

Family derives its name from the critical tyrosine residue being the fourth member of catalytic tetrad, which is highly conserved and very essential for the activity of tyrosine recombinases. Structural and site directed mutagenesis studies with tyrosine recombinases have revealed the functional role of each active site residues. Primary and secondary arginines are primarily involved in the stabilization of the transition state intermediate during crossover while histidine/tyrosine at the second catalytic position is involved in general acid base catalysis along with assistance in the transition state stabilization. The terminal tyrosine functions as nucleophile to form phospho-tyrosine intermediate with the sessile phosphate, a step very critical for the activity of tyrosine recombinases (Gibb et al., 2010).

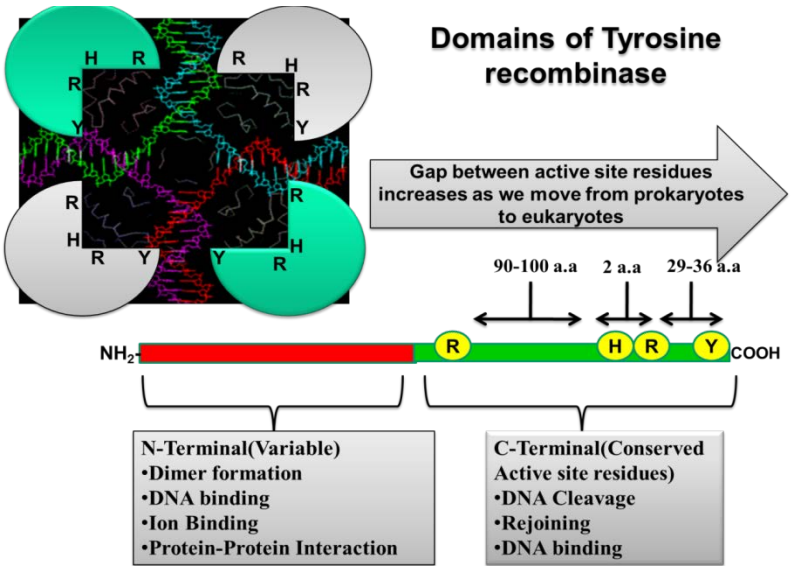

Fig. 1.11 Domain architect of tyrosine recombinase.

In addition to the tyrosine nucleophile, one lysine and two arginine residues are proposed to be essential for efficient recombination.

2.2.1.2 Serine Recombinases

Serine recombinases (SRs) carry out site-specific recombination between short DNA sequences in prokaryotes that fulfil a variety of biological functions (Smith et al., 2010). Most identified SRs are bacteriophage integrases, but some catalyze the movement of antibiotic resistance genes by transposition (Wang and Mullany, 2000; Lyras et al., 2004) and other mobile elements (Van Duyne and Rutherford, 2013), or carry out programed DNA rearrangements (Sato et al., 1990). The γδ resolvase **(Fig. 1.12a)** and Hin recombinase **(Fig. 1.12b)** are well-studied SR systems. The other functions of SRs involve resolving transposition intermediates or regulate gene expression by inversion of regulatory sequences (Grindley et al., 2006). Most resolvase/invertase enzymes have a catalytic domain of 150 residues at their amino terminus, followed by a small, helix-turn-helix (HTH) DNA-binding domain. As seen in **Fig. 1.12c,** conserved motifs of SRs reside within the catalytic domain and dimerization helix (αE) of the protein (Grindley et al., 2006). Motifs A and C contain the conserved active site residues of the recombinase. Motif D within the C-terminal portion of the E-helix is mostly involved with binding the DNA. Function of Motif B remains largely unknown despite the remarkable conservation of the Ser-39, Gly-40, and Arg-45 residues.

Process of recombination initiated with synapse formation in which four recombinase subunits get activated to attack the two crossover sites forming two double-strand breaks (**Fig. 1.12d**). This reaction covalently attaches the four recombinase monomers by a phosphor-serine linkage to the four 5’ ends of the broken strands, leaving free hydroxyls at the 3’ ends.

2.2.2 Type IB Toposiomerase

Type IB topoisomerases (TopIB) function as monomers to relax both positive and negative supercoils (Champoux, 2001b). Extensive studies were carried out with *E. coli* **(Fig. 1.13a)** and yeast topoisomerases **(Fig. 1.13b)**. Introduction of transient breaks in one strand of the DNA duplex by topoisomerase IB allows rotation about

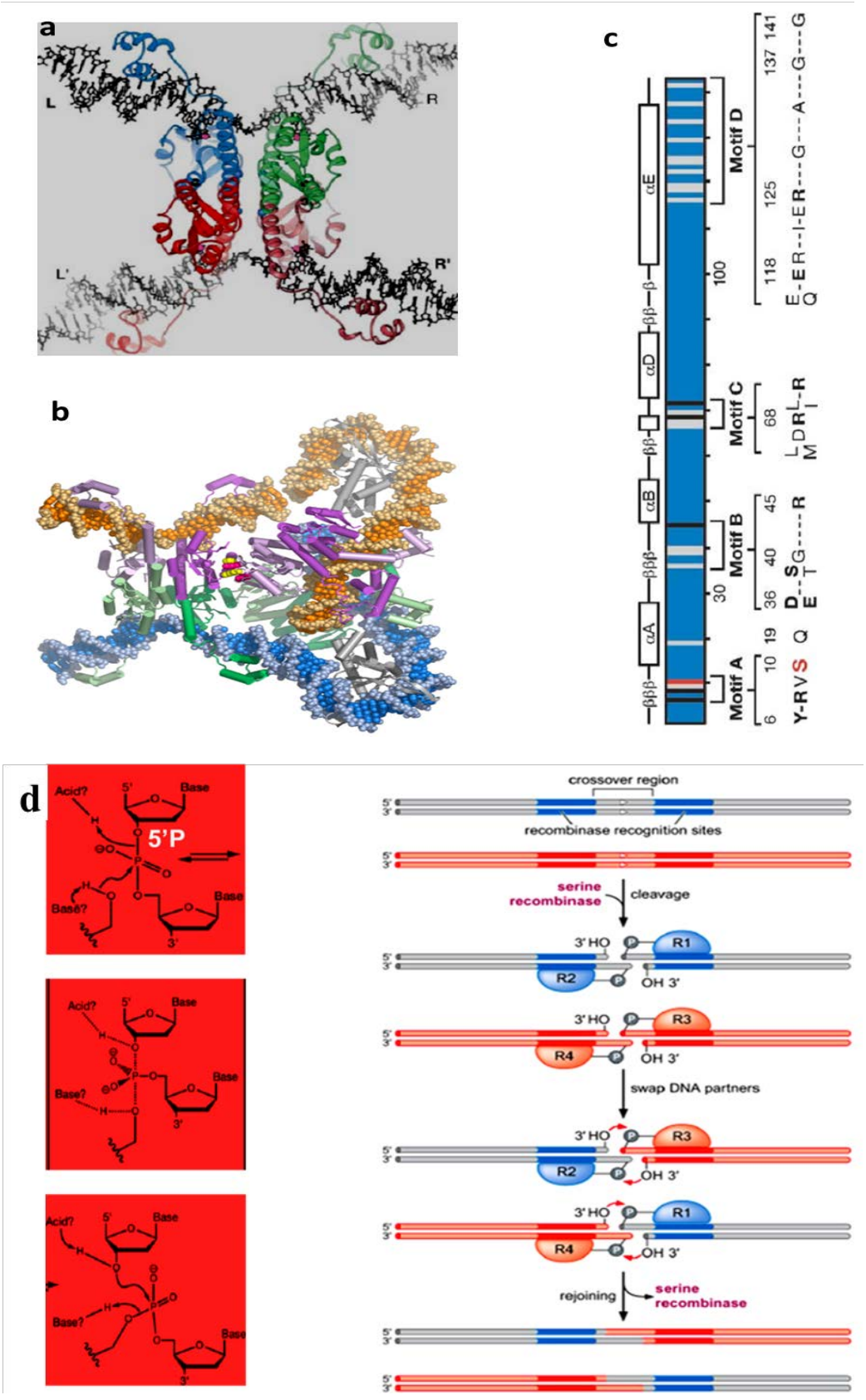

Fig. 1.12 Structure and mechanism of catalysis of Serine Recombinases (Grindley et al., 2006). (a) & (b) Crystal structure of γδ resolvase and Hin Recombinase, respectively (Grindley et al., 2006); (c) Domain architecture of serine recombinases; (d) Mechanism of catalysis of serine recombinases. Adapted and modified from Grindley et al. (2006).

the intact strand resulting in relaxation of the supercoils (Stewart and Champoux, 2001).Eukaryotic DNA topoisomerases IB modulate the topological state of DNA by cleaving and rejoining one strand of the DNA duplex (Keck and Berger, 1999). In this transesterification reaction, the scissile phosphate phosphodiester is attacked by a tyrosine of the enzyme, resulting in the formation of a DNA-(3′-phosphotyrosyl)-enzyme intermediate and the expulsion of a 5′-OH polynucleotide. During strand rejoining, the 5′-OH end attacks the covalent intermediate to form a phosphodiester bond and expel the tyrosine. There is a striking similarity between topoisomerase IB and YRs with respect to mechanism of catalysis **(Fig. 1.14)** and conservation of active site residues **(Table 1.7)**. These enzymes have critical roles in cellular processes by relieving stress by overwinding or underwinding the DNA duplex that occurs during transcription, replication, and DNA repair.

The type IB subfamily shows no sequence or structural homology to other known topoisomerases and is functionally distinct from the type IA subfamily (Champoux, 2001b; Champoux, 2001a). The eukaryotic TopIB enzymes are quite large, varying in size from 765 to 1019 amino acids. These enzymes contain an N-terminal domain, which is dispensable for relaxation activity, but contains signals for

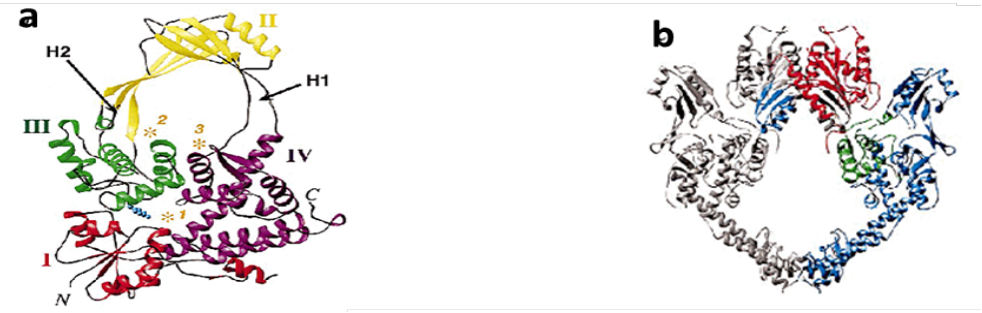

Fig. 1.13 Crystal structure of Topoisomerase IB (Keck and Berger, 1999). (a) ***E. coli*** **Topoisomerase 1.** (b) **Yeast topoisomerase IB.**

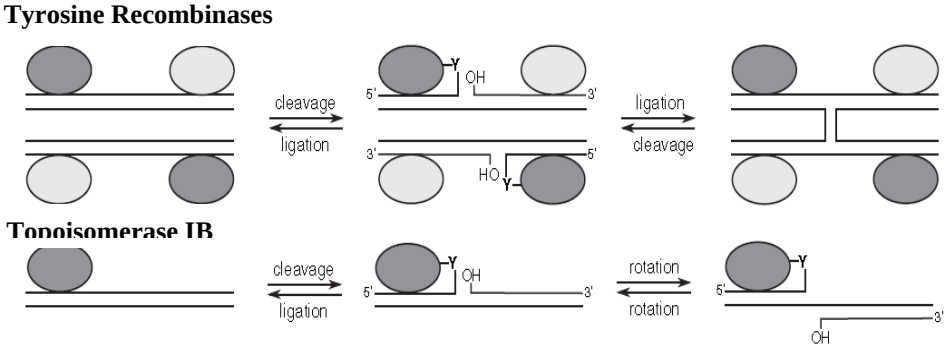

Fig. 1.14 Catalytic similarity between YRs and topoisomerase IB (Gibb et al., 2010).

nuclear localization and sites for regulatory interactions with other cellular proteins such as nucleolin, SV40, T-antigen, and a number of transcription factors (Gibb et al., 2010). Another interesting difference between eukaryotic and viral TopIB enzymes is that the poxvirus family exhibits a high degree of sequence specificity that has made the system amenable to detailed kinetic analysis.

	Arg I	Glu/Asp I	Lys	His II	Arg II	His/Trp III	Tyr
P1 Cre	R173	E176	K201	H289	R292	W315	Y324
λ Int	R212	D215	K235	H308	R311	H333	Y342
S. cerevisiae Flp	R191	D194	K223	H305	R308	W330	Y343
E. coli Xer C/D	R148	D151	K172	H240/244	R243/247	H243/247	Y275/279
Poxvirus TopIB[a]	R130	None	K167	None[b]	R223	H265	Y274
Conservation[c]	91.2% R 1.1% K	79% E 16.4% D	94.7% K 3.8% R	87.3% H 6.8% Y	99.1% R 0.5% K	92% H 4% W	99.6% Y
Functions[d]	TS	S	GAB	TS, GAB	TS	S, TS	N

Table 1.7 Conservation of active site residues between members of YRs and topoisomerase IB (Gibb et al., 2010). [a]Residue numbers refer to the vaccinia/variola virus topoisomerases. [b]A lysine residue is located in this position of the sequence in the TopIB enzymes, but is directed away from the active site. [c]The values given are based on analysis of PSI-BLAST hits from the NCBI RefSeq database using the Cre catalytic domain as query; the percentages for the catalytic lysine are estimates. [d]Roles of active site residues in catalysis, based on published biochemical and structural data. TS, transition state stabilization; S, structural; GAB, general acid/base catalysis; N, nucleophile.

2.2.3 Restriction endonucleases

Restriction endonucleases (REases) are found ubiquitously in prokaryotic system with a principle involvement in protection of host from foreign DNA, in particular bacteriophage. Other biological functions serves to involve in processes like recombination and transposition (Pingoud and Jeltsch, 2001).

Restriction endonucleases are classified mainly into four different types based on their subunit composition, sequence recognition, cleavage position, cofactor requirements, and substrate specificity (Roberts et al., 2003).

Type I enzymes consist of a hetero-oligomeric protein complex encompassing both restriction and modification activities. These enzymes bind to a bipartite DNA sequence and cleave from ~100 bp to tens of thousands of base pairs away from the target (Roberts et al., 2003). Typical examples are *Eco*KI and *Eco*R124I (Vasu and Nagaraja, 2013).

Type II REases are homodimeric or homotetrameric and cleave DNA within or near their target site. These enzymes are highly diverse and are known to utilize at least five types of folds: PD-(D/E)XK, PLD, HNH, GIY-YIG, and half pipe (Ibryashkina et al., 2007; Miyazono et al., 2007), Type II enzymes are the most widely studied and are also extensively utilized nucleases in genetic engineering.

Type III enzymes are heterotrimers (M_2R_1) (Wyszomirski et al., 2012) or heterotetramers (M_2R_2) (Vasu and Nagaraja, 2013) containing restriction-, methylation-, and DNA-dependent NTPase activities. These enzymes recognize short asymmetric sequences of 5 to 6 bp, translocate along DNA, and cleave the 3′ side of the target site at a distance of ~25 bp. Restriction is elicited only when two recognition sequences are in an inverse orientation with respect to each other.

The type IV systems cleave only DNA substrates containing methylated, hydroxymethylated, or glucosyl-hydroxymethylated bases at specific sequences (Roberts et al., 2003). The recently discovered type IV enzyme GmrSD specifically digests DNAs containing sugar-modified hydroxymethylated cytosine (Bair and Black, 2007).

2.2.3.1 Type II Restriction endonucleases (REases)

The Type II REases recognize specific DNA sequences and cleave at constant positions at or close to that sequence to produce 5′-phosphates and 3′-hydroxyls (Vasu and Nagaraja, 2013). Usually they require Mg^{2+} ions as a cofactor. They may act as monomers, dimers or even tetramers and usually act independently of their companion MTase. Because of the interest in these Type II REases for recombinant DNA technology, more than 3500 have been characterized (Roberts et al., 2003). Subtypes of Type II REases with examples are summarized in **Table 1.8.**

Subtype	Characteristic feature	Example[a]	Recognition sequence[b]
Orthodox	Palindromic recognition site, which is recognized by a homodimeric enzyme, cleavage occurs within or adjacent to the recognition site	*Eco***RI**	G↓ A A T T C C T T A A↑ G
		*Eco***RV**	G A T↓ A T C C T A↑ T A G
		*Bgl***I**	G C C N N N N↓ N G G C C G G N↑ N N N N C C G
Type IIS	Asymmetric recognition site with cleavage occuring at a defined distance	*Fok***I**	G G A T G N_9↓ N N N N C C T A C N_9 N N N N↑
Type IIE	Two sites required for cleavage, one serving as allosteric effector	*Nae***I**	G C G↓ C G C C G C↑ G C G
Type IIF	Two sites required for cleavage, both sites are cleaved in a concerted reaction by a homotetrameric enzyme	*Ngo***MIV**	G↓ C C G G C C G G C C↑ G
Type IIT	Different subunits with restriction and modification activity	*Bpu*10I	C C↓ T N A G C G G A N T↑ C G
		*Bsl***I**	C C N N N N N↓ N N G G G G N N↑ N N N N N C C
Type IIG	One polypeptide chain with restriction and modification activity	*Eco*57I	C T G A A G N_{14} N N↓ G A C T T C N_{14}↑N N
Type IIB	Cleavage on both sides of the recognition site	*Bcg*I	N N↓ N_{10} C G A N_6 T G C N_{10} N N↓ ↑N N N_{10} G C T N_6 A C G N_{10}↑N N
		*Bpl*I	N N_4↓ N_8 G A G N_5 C T C N_8 N_4 N↓ ↑N N_4 N_8 C T C N_5 G A G N_8↑ N_4 N
Type IIM	Methylated recognition site	*Dpn*I	G ^{m}A↓ T C C T↑ ^{m}A G

Table 1.8 Classification of Type II restriction endonucleases

Multiple sequence alignment of Type II REases reveals a catalytic sequence motif $PDX_{10–30}(E/D)XK$ common to majority of family members (Tamulaitis et al., 2002). The active site core comprises of 3-5 β-sheets flanked by α-helices **(Fig. 1.15)** (Pingoud et al., 2005). The second and third strand of the β-sheet serve as a scaffold for the catalytic residues of the $PDX_{10–30}(E/D)XK$ motif.

2.2.3.2 Type IIE Restriction Endonuclease

Type IIE REases share distinguishing feature of requiring cooperativity between two recognition sites in their substrate DNA (Reuter et al., 1998).

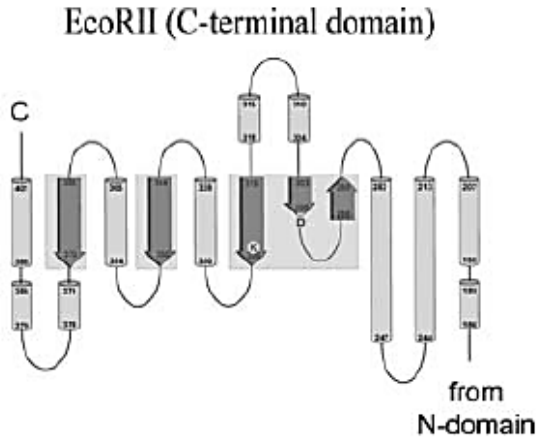

Fig. 1.15 Catalytic domain of Type II restriction endonuclease (*Eco*RII) (Pingoud et al., 2005).

They are required to interact with two copies of their recognition sequence for efficient cleavage in which one copy being the target for cleavage, the other acts as an allosteric effector (Pingoud et al., 2005). The best-studied examples of Type IIE restriction modification systems are *Eco*RII (Zhou et al., 2003) (**Fig. 1.16a)** and *Nae*I (Colandene and Topal, 1998) **(Fig. 1.16b).**

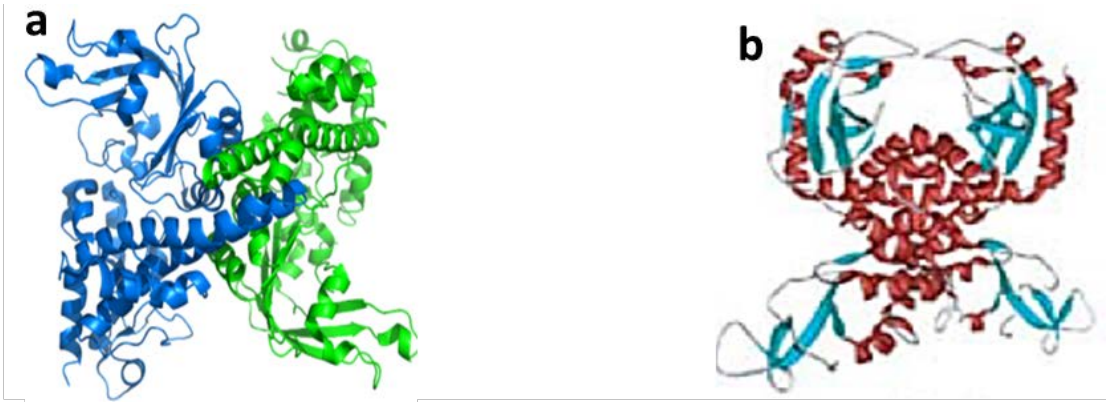

Fig. 1.16 Crystal structures of Type IIE restriction endonucleases (Pingoud et al., 2005). (a) ***Eco*RII.** (b) ***Nae*I.**

2.3 Link between Endonucleases and Site Specific Recombinases

*Eco*RII (402 aa) and *Nae*I (317 aa) are member of Type IIE REases which requires two copies of its recognition sequence for DNA cleavage (Reuter et al., 1998;

Pingoud and Jeltsch, 2001; Mucke et al., 2003; Roberts et al., 2003; Szczepek et al., 2009). The two target DNA sites may be located on the same (*in cis*) or on different DNA substrates (*in trans*) (Pingoud et al., 2005; Szczepek et al., 2009). Multiple sequence alignment studies revealed significant similarity between *Eco*RII and other members of tyrosine recombinases with P4 integrase being the closest displaying 54 % similarity and 29 % identity (**Table 9)** (Topal and Conrad, 1993).

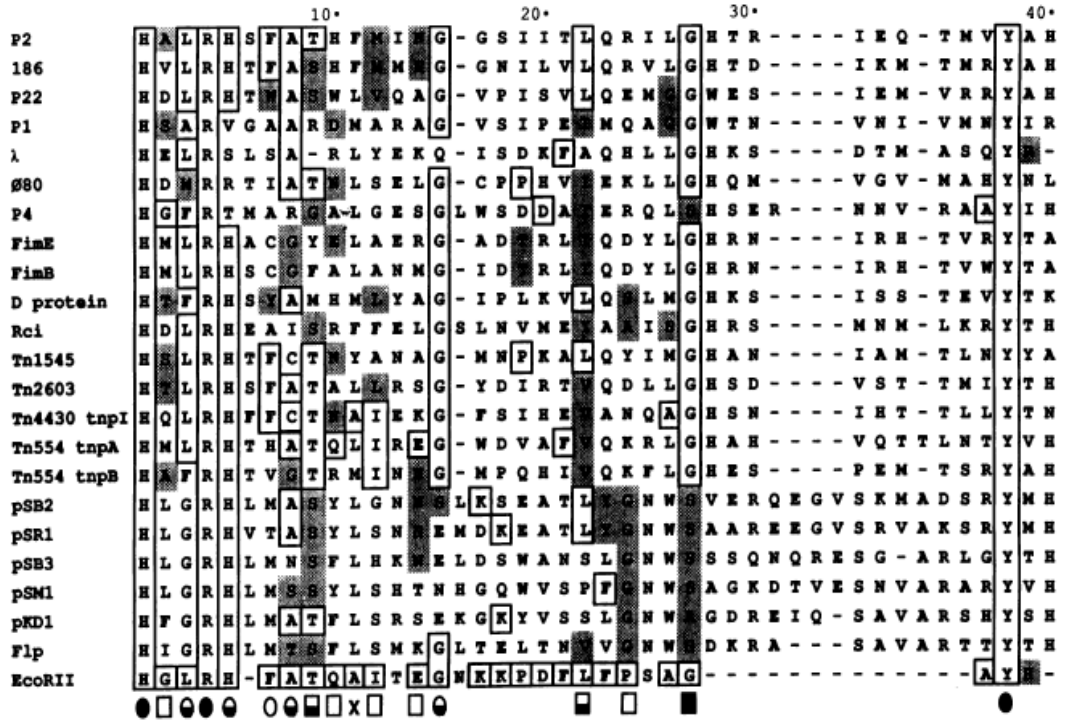

Table 1.9 Comparison between amino acid sequence of active site core of *Eco*RII and other members of tyrosine recombinase family (Topal and Conrad, 1993).

Substituting *Eco*RII Tyrosine 308 to Phenylalanine resulted in loss of cleavage activity of the enzyme, a phenomenon identical to terminal active site tyrosine of tyrosine recombinases (Topal and Conrad, 1993). *Nae*I Leucine 43 to Lysine substitution resulted in overall structural modification in the enzyme resulting in acquired topoisomerase activity of the enzyme (Jo and Topal, 1996).

XisA (472 aa) of *Anabaena* PCC 7120 is a member of tyrosine recombinase family, functionally involved in site specific excision of *nifD* element during heterocyst formation (Lammers et al., 1986). XisA is reported to display *nifD* excision in *E. coli* heterologous host despite of physiological difference with *Anabaena* PCC 7120 (Lammers et al., 1990; Karunakaran et al., 2008). XisA along with recombinase

also possesses restriction endonuclease activity at its 11 bp target site (**C*GGA*GTAA*TCC***) (Shah et al., 2007). XisA requires two target sites to display endonuclease activity, a mechanism similar to *Eco*RII and *Nae*I suggests that XisA belongs to Type IIE class of REase. Involvement of 308 Tyrosine of *Eco*RII integrase domain (Topal and Conrad, 1993) in endonuclease activity suggests functional similarity between XisA and *Eco*RII which is suggestive of XisA being an evolutionary intermediate between recombinases and endonuclease.

Even though the *xisA* gene was cloned and sequenced long ago (Lammers et al., 1986), neither the protein nor the transcript has been detected so far in *Anabaena* or in *E. coli* strains expressing *xisA.* It is therefore necessary to overexpress and purify XisA protein using heterologous system. Earlier studies in our laboratory established XisA endonuclease activity in addition to recombinase activity (Shah et al., 2007). The relationship between the structural domains of the XisA to these two activities is not known. Thus, it is necessary to decipher the recombinase and endonuclease activities in relation to N- and C-terminal regions of XisA.

Based on the background information, objective of the present study are as follows:

2.4 Objectives

2.4.1 Expression and purification of functional XisA protein

2.4.2 Development of PCR based assay to monitor the recombinase activity of XisA protein *in vivo*

2.4.3 Determining minimal regions of XisA protein displaying recombinase and endonuclease activities

2.4.4 Role of predicted basic region of leucine zipper of XisA protein in recombinase and endonuclease activities

2.4.5 Role of active site residues in recombinase and endonuclease activities of XisA protein