Extracytoplasmic function sigma factors in Pseudomonas aeruginosa


To cite this version:

HAL Id: hal-01788719
https://hal.science/hal-01788719
Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License
Transcriptional regulation of development in heterocyst-forming cyanobacteria

Enrique Flores*, Silvia Picossi, Ana Valladares and Antonia Herrero

Instituto de Bioquímica Vegetal y Fotosíntesis, CSIC and Universidad de Sevilla, Seville, Spain

*E-mail address: eflores@ibvf.csic.es

ABSTRACT

Filamentous, heterocyst-forming cyanobacteria are among the simplest multicellular systems in Nature. In the absence of combined nitrogen, the filaments consist of vegetative cells that fix CO₂ through oxygenic photosynthesis and micro-oxic heterocysts specialized for the fixation of N₂ in a proportion of about 10 to 1. The development of a heterocyst-containing filament involves differentiation of vegetative cells into heterocysts in a process that requires a distinct gene expression program. Two transcription factors are strictly required, the CRP-family NtcA and HetR. NtcA directly activates the expression of multiple genes during heterocyst differentiation—in some cases assisted by coactivators including HetR—and in mature heterocysts, whereas HetR is needed to build high NtcA levels in differentiating heterocysts and directly activates some particular genes. A few other regulators of gene expression participate at specific differentiation steps, and a specific transcription factor, CnfR, activates nif gene expression under the micro-oxic conditions of the heterocyst.
1. Introduction

Multicellularity is a form of organization widely found in living organisms including evidently eukaryotes but also prokaryotes. Two properties characterize multicellularity: cell-cell binding and cell-cell communication. Additionally, organisms take advantage of multicellularity to specialize cells in different functions, so that cellular differentiation is another hallmark of multicellularity. Proterozoic microfossils frequently show a multicellular form, and filamentous cyanobacteria showing cellular differentiation have been traced back to about 2.45 – 2.1 billion years ago [1]. Cyanobacteria are a phylogenetically coherent group of organisms characterized by performing oxygenic photosynthesis, which allows a mainly photoautotrophic mode of growth [2]. Additionally, many cyanobacteria can fix atmospheric N\textsubscript{2}, making these organisms capable to grow with very simple nutritional sources –air and water with a few mineral salts. However, because the N\textsubscript{2}-fixation machinery is very sensitive to oxygen [3], cyanobacteria need to separate N\textsubscript{2} fixation from oxygenic photosynthesis. Whereas some unicellular and filamentous cyanobacteria separate these two incompatible processes temporally, restricting N\textsubscript{2} fixation to the dark period of diel cycles [3], some filamentous cyanobacteria form cells specialized for N\textsubscript{2} fixation. Thus, in filaments of cells that fix CO\textsubscript{2} through oxygenic photosynthesis, some of these cells (usually less than 10 \%) differentiate into \textit{N}_2-fixing heterocysts when no combined nitrogen is available [4]. Other developmental options found in some heterocyst-forming cyanobacteria include the differentiation of akinetes, a type of spores that resist drought and cold, and the production of hormogonia, short motile filaments frequently made of small-sized cells that serve a dispersal function [5, 6]. Knowledge of the transcriptional regulation for the differentiation of akinetes and hormogonia is however scarce [6]. This review will focus on the transcriptional regulatory mechanisms that operate in heterocyst differentiation. We shall first briefly introduce the structure of the cyanobacterial filament and the morphology and biochemistry of the heterocyst, then describe the mechanisms of regulation of gene expression that operate during the process of differentiation and in the mature N\textsubscript{2}-fixing heterocyst, and finally discuss the early regulation that leads to a defined pattern of heterocyst distribution in the filament.

2. Cyanobacterial filaments and heterocysts

Cyanobacteria of the order Nostocales –taxonomic section IV in the classification of Rippka \textit{et al.} [2]– are characterized by their ability to produce heterocysts when they lack a source of combined nitrogen [2]. These cyanobacteria grow as filaments that can be hundreds of cells
long, and heterocysts can be found intercalary or at the end in the filaments, although some strains make only terminal heterocysts [2]. Heterocysts are terminally differentiated cells, and commitment to complete differentiation is established after the spatial distribution pattern of heterocysts has already been settled [7, 8].

2.1. Structure of the cyanobacterial filament
The cyanobacteria are diderm bacteria, i.e., they bear a Gram-negative type of cell envelope. The cells in a cyanobacterial filament are surrounded by their cytoplasmic membrane and peptidoglycan layer(s), but the outer membrane does not enter into the septa between consecutive cells and, therefore, is continuous along the filament [9]. This implies that the periplasm –the space that lies between the cytoplasmic and outer membranes and contains the peptidoglycan– is also continuous and may be a communication conduit between the cells in the filament [10]. Consecutive cells in the filament are also joined by proteinaceous structures known as septal junctions [9, 11], which appear to traverse the septal peptidoglycan through perforations known as nanopores [12, 13]. These perforations are drilled in the peptidoglycan by AmiC-type amidases [12, 14]. In the model heterocyst-forming strain Anabaena sp. PCC 7120 (hereafter Anabaena), the SepJ (a.k.a. FraG), FraC and FraD proteins are localized at the cell poles in the intercellular septa of the filament [15, 16]. They are membrane proteins that, in the case of SepJ and FraD, have long extra-membrane sections that appear to be located outside of the cells in the intercellular septa, and at least SepJ is known to form multimers [17, 18]. Hence, they are candidate components of the septal junctions (see below).

2.2. Heterocyst morphology
Under the light microscope, the heterocysts are seen larger than the vegetative cells and show two refractile bodies at their poles (just one, proximal to the vegetative cells, in the case of terminal heterocysts) (Fig. 1A). Under the electron microscope, differences become evident in the global form of the heterocyst, in its envelope, and in its cytoplasm (Fig. 1B). The heterocyst poles adjacent to vegetative cells take a narrow form known as “neck”, which makes the area of contact between heterocysts and vegetative cells smaller than that between vegetative cells. The cell envelope contains extra material that is deposited outside the outer membrane and consists of an inner glycolipid (heterocyst-specific glycolipid, Hgl) layer and an outer polysaccharide (heterocyst envelope polysaccharide, Hep) layer. Three characteristics distinguish the heterocyst cytoplasm from the vegetative cell cytoplasm [6]: (i) intracellular membranes are reorganized forming a structure known as “honeycomb” around the
heterocyst poles; (ii) glycogen granules and carboxysomes (the intracellular micro-
compartments that contain ribulose bisphosphate carboxylase/oxygenase [RuBisCO, the CO$_2$-
fixing enzyme]) disappear; and (iii) a large granule made of cyanophycin (multi-L-arginyl-poly-
L-aspartic acid]) is deposited at the heterocyst poles adjacent to vegetative cells –these
granules correspond to the refractile bodies observed by light microscopy.

2.3. Heterocyst biochemistry

The heterocyst has evolved to become a N$_2$-fixing factory. Hence, an essential goal of its
biochemistry is to keep a micro-oxic environment appropriate for the expression and function
of the N$_2$-fixing enzyme, nitrogenase, and accessory proteins. First, importantly, the oxygen-
releasing photosystem II (PSII) is non-functional in the heterocyst (reviewed in [19]). Second,
the extra envelope described above restricts gas diffusion contributing to reduce the amount
of oxygen that enters in the heterocyst as air (78% N$_2$, 21% O$_2$) [20]; indeed, O$_2$ has been
suggested to enter the heterocyst mainly through the vegetative cell-heterocyst connections
[20]. Third, dedicated terminal respiratory oxidases located in the honeycomb membranes and
cytoplasmic flavodiiron proteins consume O$_2$ contributing to create a micro-oxic environment
[21, 22]. The nitrogenase reaction consumes high amounts of ATP and reductant, and the
heterocyst keeps an electron transport chain and PSI activity that support
photophosphorylation and the reduction of ferredoxin that can provide nitrogenase with
electrons [19].

In the heterocyst, dinitrogen is reduced by the nitrogenase complex producing
ammonium (and hydrogen as a byproduct). The ammonium resulting from the N$_2$-fixing
reaction is incorporated by glutamine synthetase into glutamate producing glutamine [23].
Glutamine is, at least in part, transferred to vegetative cells in exchange for glutamate, which
is not synthesized at high levels in the heterocyst [24, 25]. Further anabolic nitrogen
metabolism in the heterocyst includes the biosynthesis of aspartate and arginine, which are
incorporated by cyanophycin synthetase into cyanophycin that serves as a dynamic nitrogen
reservoir [26, 27]. In addition to glutamine, β-aspartyl arginine, the product of cyanophycin
degradation by cyanophycinase, is transferred to vegetative cells, in which β-aspartyl arginine
is hydrolyzed by isoaspartyl dipeptidase releasing aspartate and arginine [28]. Hence, both
glutamine and β-aspartyl arginine serve as nitrogenous nutrients for the vegetative cells. In
turn, because the heterocyst does not accomplish the photosynthetic fixation of CO$_2$, the
vegetative cells provide the heterocysts with reduced carbon in the form of a sugar, sucrose,
that in *Anabaena* is split by an invertase –InvB– producing glucose and fructose, which
support heterocyst metabolism [29, 30]. This exchange of reduced carbon by fixed nitrogen is at the basis of the multicellular behavior of heterocyst-forming cyanobacteria.

2.4. Intercellular molecular exchange

During heterocyst differentiation and in the diazotrophic filament, regulators and metabolites are transferred between cells. Intercellular molecular exchange in filamentous cyanobacteria has been probed with fluorescent markers including calcein, 5-carboxyfluorescein and the sucrose analog esculin [13, 31, 32]. The transfer has properties of simple diffusion, indicating the existence of direct connections between adjacent cells [31, 32]. Mutants of Anabaena lacking the septal proteins SepJ or FraC and FraD are impaired in the intercellular transfer of those fluorescent markers and make less nanopores than the wild type, suggesting that these proteins contribute to the formation of the septal junctions [13, 31]. Differences in the impairment of transfer of the various fluorescent markers in different mutants have suggested the existence of functionally-discernible SepJ- and FraCD-related septal junctions [13, 17]. Important for our discussion here is that septal junction complexes in the vegetative cell/heterocyst septa may have specific characteristics as compared to those in vegetative cell/vegetative cell septa. Thus, these structures appear to be longer in the former than in the latter [33], and the SepJ protein visualized as a SepJ-GFP fusion is seen as a single fluorescent spot between vegetative cells and as a double spot in the vegetative cell/heterocyst septa, implying a more spread localization in the latter than in the former [11].

3. Key transcription factors and transcriptional regulation mechanisms

Two genes whose inactivation completely abolish heterocyst differentiation are ntcA [34, 35] and hetR [36]. Whereas NtcA is a global regulatory transcription factor of cyanobacteria mediating responses to variations in the C-to-N balance of the cells, HetR is a transcription factor specifically found in filamentous cyanobacteria whose best-known role is in cell differentiation.

3.1. NtcA

NtcA is a cyanobacterial transcription factor that orchestrates the genetic responses to the C-to-N balance by directly regulating the expression of multiple genes involved in pathways for nitrogen assimilation, but also of a number of other metabolic pathways [37, 38]. Indeed, by means of chromatin-immunoprecipitation assays performed in Anabaena, an extraordinarily
large number of NtcA-binding sites on DNA were identified that could be ascribed to more than 2,000 genes [38]. Besides genes involved in nitrogen scavenging and N₂ fixation, the NtcA targets included genes involved in DNA metabolism, transcription and translation, central metabolism and, remarkably, many predicted regulatory genes, which should extend the range of NtcA-influenced cellular processes. NtcA belongs to the CRP (cAMP receptor protein) family of transcriptional regulators, presenting a domain architecture similar to that of the well-characterized *E. coli* CRP protein, including an N-terminal β-roll fold, which in CRP accommodates the effector cAMP, a central long dimerization helix and a C-terminal helix-turn-helix (HTH) motif for interaction with DNA [39] (Fig. 2A). These regulators are dimeric proteins, so that each subunit interacts with half of a palindromic recognition sequence in DNA.

NtcA can act as a transcriptional activator or repressor. For both activities, the NtcA dimer binds to palindromic DNA sites including the consensus sequence GTAN₈TAC [37, 38]. In many NtcA-activated promoters an NtcA-binding site is found separated by ca. 22 bp from a -10 promoter determinant of consensus sequence TAN₅T, thus conforming to the so-called Class II bacterial activated promoters, in which the regulator binds in place of a missing or degenerated -35 determinant. This promoter structure is frequently found in genes of unicellular cyanobacteria and, in heterocyst-forming strains, in vegetative cell-expressed genes as well as in genes expressed in mature heterocysts [40]. Besides this, NtcA can activate gene expression from DNA sites centered further upstream the -35 position, conforming to Class I activated promoters, or from promoters with very degenerated NtcA-binding sites, which are generally activated during heterocyst differentiation [41]. Finally, repression by NtcA has been shown involving binding sites downstream from promoter determinants, overlapping the -10 box or even within the regulated ORF [41].

Regarding metabolic signals, NtcA activity is influenced by 2-oxoglutarate (2-OG) [42]. Because cyanobacteria lack the enzyme 2-oxoglutarate dehydrogenase, the 2-OG cellular levels depend on its synthesis during CO₂ fixation and its consumption mainly to form glutamate by the glutamine synthetase/glutamate synthase pathway, thus representing an indicator of the cellular C-to-N balance [40]. Indeed, 2-OG has been shown to increase the binding of NtcA to DNA in many regulated promoters *in vitro*, although in contrast to CRP, NtcA can interact with DNA in the absence of the effector (see e.g. [37, 42, 43]). Structurally, the positive effect of 2-OG in binding is explained by its effect on dimer stabilization by decreasing the distance between the dimerization helices of each subunit, which in turn repositions the two second helices in the HTH DNA-binding domains, making the global conformation more suitable for interaction with DNA [39]. In the absence of the effector, the
NtcA apo-protein already presents in its DNA-binding domain some exposed positively-charged residues that represent a DNA-binding patch. In addition, the conformation of the last helices in this domain allows some interaction with DNA, whereas these helices are buried in the apo-CRP [39].

Whereas a positive effect of 2-OG on NtcA binding to DNA in vitro has been found in all types of NtcA-activated promoters, although at variable extents, an additional and stringent requirement for 2-OG has been found for the step of promoter melting to form the tripartite NtcA-RNA polymerase-DNA open promoter complex, and thus for transcript production, at Class II NtcA-activated promoters [43]. However, no specific structural information is available for this 2-OG effect, although it is reminiscent of that of cAMP during Class II activation by CRP. Finally, the intrinsic capacity of NtcA to interact with DNA may be important for its action as a repressor under physiological conditions that determine low 2-OG levels. This appears to occur in the cmpR gene (encoding a LysR-type regulator of a bicarbonate transporter operon), which under conditions that determine low 2-OG levels is expressed at higher levels in an ntcA mutant than in the wild type [44].

3.2. HetR

For many years considered as a heterocyst-specific regulator, evidence is recently accumulating in support of HetR activity also in undifferentiated cells, consistent with the presence of HetR in non-heterocystous cyanobacteria as well as in heterocyst formers [45]. HetR is required for the expression of multiple genes involved in heterocyst differentiation, and binding of HetR to DNA from the promoter region of a few of them has been detected in vitro, including sequences from the promoters of hetR, patS and hepA ([46], see below) as well as of hetP, which has been implicated in the commitment to irreversible differentiation [47, 48]. A global study of HetR targets by chromatin immunoprecipitation identified 26 direct targets, of which only 10 % are related to known heterocyst-differentiation genes [49]. Finally, a role of HetR in repression of several gene promoters in vegetative cells has been reported [29, 50, 51].

Structurally, HetR represents a rather distinct type of transcription factor. The crystal structure of a dimer of HetR from the heterocyst-forming cyanobacterium Fischerella sp. shows a central DNA-binding region constituted by the N-terminal HTH motifs of both subunits, and new folds conforming two globular "flaps" and a "hood" over the central core, the latter formed by the two C-terminal parts [52] (Fig. 2B). However, HetR tetramers have been detected in vivo, at higher levels during heterocyst differentiation, leading to the proposal that the tetramer represents an active form of HetR [53]. This is consistent with (i)
the fact that in crystals assembled by HetR in complex with long synthetic DNA targets, HetR
tetramers formed by interaction of the flap domains were the predominant form [54], and (ii)
the occurrence of tandem repeats of the HetR DNA-binding consensus sequence in several of
the identified HetR targets [49].

3.3. Mechanisms of transcriptional activation by NtcA and HetR

The expression of both ntcA and hetR increase early during heterocyst differentiation, upon
combined N step-down, in a mutually-dependent manner enhanced by positive
autoregulation, with hetR induction preceding that of ntcA [55]. Remarkably, induction takes
place in specific cells or cell clusters along the filament, marking the places of vegetative cell
differentiation into heterocysts ([56-58], see below). The response regulator-like factor NrrA,
which is directly induced by NtcA early upon N step-down from a canonical Class II promoter
[59], has been reported to bind the hetR promoter region in vitro and, hence, to mediate the
induction of hetR by NtcA [60].

Multiple genes whose products effect the morphological and biochemical
differentiation of heterocysts (see below), including secondary regulators, are also activated
upon combined N step-down at specific steps during differentiation [41, 61], and for some of
them induction could be ascribed to specific cells along the filament (see e.g. [29, 62]). In
contrast to nrrA, which is independent of HetR, in vivo activation of many other heterocyst-
differentiation genes depends on both NtcA and HetR, as they are not induced in strains
bearing inactivated ntcA or hetR genes. Mechanistically, a remarkable feature of heterocyst-
differentiation genes is that they present complex promoter sequences including several
successive promoters, which may include σ70 consensus-type promoters, likely directing a
basal level of gene expression in all the cells of the filament before N step-down, Class II and
Class I NtcA-activated promoters, and HetR-dependent promoters (see e.g. [41, 50, 63]). The
binding affinity of NtcA for the different NtcA-activated promoters and the response to HetR
are important in determining the temporal sequence of transcription activation during the
differentiation process that, as a result of the cell-confined increase of hetR and ntcA
expression, takes place in specific cells along the filament.

HetR appears to exert a positive effect on transcription activation by several
mechanisms. As mentioned above, direct binding of HetR in vitro has been reported for only a
few sequences upstream of heterocyst-differentiation genes [49], suggesting that the in vivo
requirement could be indirect in many promoters, e.g., by the requirement of high levels of
NtcA, attained by HetR-dependent activation of the ntcA gene, or of later secondary
regulators. In the case of the proximal promoter of the devBCA operon involved in deposition of the Hgl layer (see below), HetR could not bind DNA by itself but was found to increase the binding of NtcA at degenerated NtcA-binding sequences, thus behaving as an NtcA-coactivator [50] (Fig. 2C). On the other hand, the pipX gene, which is activated directly by NtcA at late stages of heterocyst differentiation, encodes a small protein required for full diazotrophic growth and full activation of further late genes [62]. The structure of a complex of the NtcA and PipX proteins from the unicellular cyanobacterium Synechococcus elongatus has been determined, leading to the suggestion that binding of PipX stabilizes the active conformation of NtcA [64]. With the Anabaena proteins, PipX has been shown to increase NtcA binding to DNA and transcription from NtcA-dependent promoters in vitro, thus representing a late NtcA-coactivator [65].

Regarding RNA polymerase sigma factors, NtcA-dependent transcription has been achieved in vitro with the Anabaena RNA polymerase holoenzyme including the principal group 1 σ\textsuperscript{70}-family factor SigA [43]. In Anabaena, the sigA gene is transcribed from one major, proximal promoter that was first described to display constant activity [66] and later to be repressed by NtcA after N step-down [67]. Additionally, several promoters induced upon removal of combined nitrogen contribute to expression of sigA. This array of promoters produces increased transcription of sigA in the differentiating cells, which is consistent with a bulk of NtcA-dependent transcription during differentiation. Further, Anabaena has multiple σ\textsuperscript{70}-family factors of groups 2 and 3 [68]. Although inactivation of none of them abolishes heterocyst differentiation or diazotrophic growth, some of the single and double mutants that have been isolated are affected in these processes [69]. Also, the sigC (group 2), sigE (group 2) and sigG (group 3) genes are responsive to N step-down, with successive increases of expression in the cells that are differentiating into heterocysts [70]. SigE is upregulated late during differentiation, and both NrrA [71] and NtcA [72] directly bind to sequences upstream of sigE. SigE is required for normal cellular differentiation and for expression of several heterocyst-specific genes including nifH (encoding a nitrogenase subunit), fdxH (encoding a putative electron donor to nitrogenase [73]), hglE\textsubscript{2} (encoding a heterocyst-specific glycolipid biosynthesis protein [72]), and the xfp operon involved in glycolysis [71]. Thus, in spite of considerable redundancy, a succession of specific σ\textsuperscript{70}-family factors appears to cooperate with SigA during heterocyst differentiation.

Finally, HetR is a target of phosphorylation in vitro and phosphorylation interferes with the formation of HetR tetramers and HetR binding to DNA [53]. Consistently, overexpression in Anabaena of a heterocyst-specific kinase, PknE, blocks heterocyst differentiation in a process
that appears to involve inactivation of HetR [74]. The pknE gene is induced late during differentiation [61] with a direct participation of HetR [74] and NtcA [38], the latter using a putative Class II promoter. These independent, in vitro and in vivo observations strongly suggest that phosphorylation of HetR is a physiological process, leading to the proposal that phosphorylation represents a mechanism of HetR inactivation at late stages of heterocyst differentiation, when a direct action of HetR would be no longer required [53, 74].

4. Regulation of gene expression in the course of heterocyst differentiation

The expression of genes involved in heterocyst differentiation and function is generally dependent on NtcA and HetR. Although regulation by these transcription factors may in some cases be indirect, evidence for direct regulation, especially by NtcA, is also available for many genes. Additionally, a number of other regulators, some of them transcription factors, have been shown to be involved in the expression of genes at specific stages of heterocyst differentiation.

4.1. Biosynthesis and deposition of the heterocyst envelope

Key steps in the morphological and metabolic differentiation of heterocysts are the deposition of the Hep and Hgl layers. During differentiation, the Hep layer is deposited before the Hgl layer, allowing the latter to lie beneath the former.

4.1.1. Hep layer

In the chromosome of Anabaena, numerous genes involved in Hep layer formation are clustered in a “gene expression island”, which covers ORFs alr2825 to alr2841 [75], and are induced under nitrogen deprivation [76]. Because they are tightly packed together, the genes alr2825 to alr2831 might constitute an operon, as might be the case for the genes alr2835 (hepA) to alr2841 as well [61]. In spite of the presence of numerous NtcA binding sites in this region [38], direct NtcA-activated expression has not been reported. On the other hand, induction in response to nitrogen deprivation of the hepA gene within this cluster requires HetR [56], which binds to a DNA fragment from the hepA promoter [46]. Direct regulation of hepA by HetR establishes a link between formation of the Hep layer and the developmental cascade of gene regulation that operates heterocyst differentiation (Fig. 3).
Expression of *hepA* is additionally dependent on HepK, a protein-histidine kinase of a two-component regulatory system that shows homology to bacterial proteins involved in acclimation to changing O$_2$ conditions [77]. Interestingly, a normal oxic atmosphere is required for differentiation of mature heterocysts [78], but it remains to be determined whether HepK actually mediates a response to O$_2$. A cognate response regulator of HepK is DevR [79], which can be phosphorylated but lacks a C-terminal effector domain (RR* in Fig. 3). This suggests that HepK/DevR participate in a phosphorelay transduction pathway that includes further protein components, in particular a possible DNA binding protein [80].

Three additional putative regulatory genes that are required for production of the Hep layer are *hepN* (*alr0117*) that encodes a histidine kinase that lacks a sensor domain (HK* in Fig. 3) [81, 82], *hepS* (*all2760*) that encodes a putatively membrane-anchored Ser-Thr kinase [82], and *henR* (*alr1086*) that encodes a response regulator containing a CheY-like receiver domain and a PPM-type (magnesium or manganese-dependent protein phosphatase that dephosphorylates Ser and Thr residues) phosphatase domain [82]. The effect of inactivation of some of these genes on global gene expression in *Anabaena* – analyzed with oligonucleotide microarrays – has suggested that HepS and HenR belong to the same regulatory system, and that HepN is part of the HepK/DevR regulatory pathway as well as of another undetermined pathway [83]. To which environmental cues the HepS kinase and the HenR phosphatase respond is unknown, although expression of *hepS* is activated under nitrogen deprivation [61].

In summary, genes involved in formation of the Hep layer are subjected to regulation within the cascade of gene expression that operates during heterocyst differentiation promoted at least by HetR, but also to pathway-specific regulation involving two-component regulatory system(s) and protein kinases/phosphatases. A main goal of future research should be to find out whether any of these pathway-specific regulators responds to oxygen, and a point of general interest is whether any of them are subjected to direct NtcA/HetR-dependent regulation.

4.1.2. Hgl layer

For a detailed discussion of the heterocyst envelope Hgl layer, the excellent review of Awai et al [84] is recommended. As is the case for the Hep layer, numerous genes involved in the formation of the Hgl layer are clustered in the chromosome of *Anabaena* (ORFs *all5341* to *all5359*), including genes involved in the biosynthesis of the glycolipids and genes involved in their deposition outside of the outer membrane [85, 86]. Further genes necessary for the formation of the Hgl layer are located outside this cluster, and these include the *devBCA* operon encoding and ABC exporter that, together with a TolC-like protein (HgdD), constitutes a
Type I secretion system involved in the export of glycolipids outside of the outer membrane in the differentiating heterocyst [87]. Along the Hgl cluster, NtcA binds to several sites, including sites upstream of genes *all5341* (*hglT*) and *asr5350* as well as in the region between *all5347* and *all5348*, which could constitute activator sites [38]. On the other hand, as described above, the promoter of the *devBCA* operon is a well-characterized, complex NtcA-dependent promoter [50, 88, 89]. It contains one canonical Class II NtcA-activated promoter and one non-canonical promoter that appears to be activated by NtcA assisted by HetR [50].

Inactivation of the *hepS* and *henR* genes involved in formation of the Hep layer also down-regulates expression of genes involved in the formation of the Hgl layer such as *hglE* (alt5351) [83]. On the other hand, DevH is a CRP-family transcriptional regulator whose expression is increased under nitrogen deprivation in an NtcA- and HetR-dependent mode [90]. Inactivation of *devH* affects specifically the expression of Hgl-biosynthesis genes such as *hglE* and the production of Hgl [91]. Whether HepS/HenR and DevH act in concert in the regulation of expression of *hgl* genes is unknown (discussed in [83]). *Anabaena* has two heterocyst-specific glycolipids. Production of the minor one appears to be dependent also on two Ser/Thr kinases, Pkn44 (All1625) and Pkn30 (All3691) [92], and production of the major one appears to be dependent on a specific protein phosphatase, PrpJ1 (All1731), that also affects expression of *hglE* [93]. The possibility that Pkn44/Pkn30 and PrpJ1 belong to two different regulatory branches for production of the minor and major Hgls, respectively, acting downstream of DevH has been discussed [92].

In summary, expression of genes for the formation of the Hgl layer involves transcription factors including the well-known NtcA and HetR proteins discussed above and DevH. In the case of DevH, the mechanism of dependence on NtcA and HetR and the direct DNA targets remain to be elucidated. Additionally, protein kinases and phosphatases affect the formation of the Hgl layer as a whole and specifically the major and minor glycolipids of that layer. The environmental cues to which such kinases and phosphatases respond and how they affect glycolipid production remain to be determined (e.g., they could affect transcription factors or directly the activity of biosynthetic enzymes).

### 4.2. Oxygen-consuming enzymes and protection against ROS

Oxygen that enters the heterocyst is reduced by at least two types of enzymatic processes: respiration [4] and reduction by flavodiiron proteins [94]. Additionally, a ruberythrin, RbrA, protects against oxidative stress in the heterocyst [95]. In *Anabaena*, two terminal respiratory oxidases, Cox2 and Cox3, substitute in the heterocyst for the terminal respiratory oxidase,
Cox1, expressed in vegetative cells [21]. Cox2 is an \( \alpha_2\)-type cytochrome \( c \) oxidase encoded by the \( \text{coxB2-coxA2-coxC2} \) operon and Cox3 appears to be an heme-copper quinol oxidase that is encoded by the \( \text{alr2729-alr2730-coxB3-coxA3-coxC3} \) operon. The contribution of Cox3 to protection of nitrogenase appears to be more important than that of Cox2 [21, 96]. The expression of both operons is strictly dependent on NtcA and HetR. The \( \text{cox3} \) promoter region has been characterized in detail and found to contain both Class I and Class II NtcA-dependent promoters in which NtcA appears to work assisted by PipX [65].

Flavodiiron proteins directly reduce \( \text{O}_2 \) to water without the production of deleterious reactive oxygen species (ROS) [97]. In \textit{Anabaena}, two flavodiiron proteins, Flv1B and Flv3B, are specifically expressed in the heterocysts [94], being Flv3B especially relevant for diazotrophic performance [22]. These flavodiiron proteins are the products of two contiguous genes, \( \text{all0178} \) (Flv3B) and \( \text{all0177} \) (Flv1B), which may constitute an operon expressed from an NtcA-dependent promoter [38].

Interestingly, scavenging of \( \text{O}_2 \) by terminal respiratory oxidases and flavodiiron proteins appears not to be redundant [38]. Whereas the terminal respiratory oxidases likely remove \( \text{O}_2 \) in the heterocyst poles, since respiratory activity is located at the honeycomb membranes [98], the heterocyst flavodiiron proteins are cytoplasmic [94]. Photosystem I-generated reductants can also reduce \( \text{O}_2 \), but in this case producing ROS. In \textit{Anabaena}, the ruberythrin \( \text{RbrA} \) is a heterocyst-specific peroxidase that transfers electrons from NADPH to \( \text{H}_2\text{O}_2 \) with the participation of ferrodoxin-NADP\(^+\) oxidoreductase [95]. The \( \text{rbrA} \) gene is expressed from two promoters one of which directs expression in the heterocysts, is dependent on HetR and binds HetR \textit{in vitro} [95].

4.3. Particular heterocyst metabolism genes

The metabolism of the heterocyst is focused on the fixation of \( \text{N}_2 \) and the assimilation of the \( \text{N} \) fixed in that reaction. We shall address now the transcriptional regulation of \( \text{nif} \) genes and of nitrogen assimilation genes such as those encoding glutamine synthetase and cyanophycin metabolism enzymes, as well as other relevant genes.

4.3.1. \( \text{nif} \) gene expression

In diazotrophic bacteria, the nitrogenase structural genes, \( \text{nifH-nifD-nifK} \), are generally clustered together with other nitrogen fixation genes –including, e.g., genes encoding nitrogenase electron donors such as specific ferredoxins– and nitrogenase maturation genes –such as FeMo-cofactor biosynthesis genes (99). The \textit{Anabaena} genome bears a large \( \text{nif} \) gene
cluster including nifB-fdxN-nifS-nifU-nifH-nifD-nifK-nifE-nifN-nifX and a few other genes downstream. Notably, fdxN and nifD are interrupted by long DNA elements that are excised during heterocyst differentiation [100]. After excision, the nifHDK operon is mainly expressed from a promoter located upstream from nifB [101], although the large mRNA produced is subjected to processing so that nifHDK, nifHD and nifH transcripts are produced (see e.g. [96]).

In *Anabaena* and other diazotrophic cyanobacteria, including non-heterocystous strains, the nifB promoter is activated by CnfR [102], previously known as PatB [103]. The CnfR protein has N-terminal sequences for binding two 4Fe-4S clusters, resembling bacterial-type ferredoxins, and a C-terminal HTH motif for binding to DNA [103, 104]. It is thought that CnfR can activate nif gene expression in response to a low-oxygen tension [104, 105], and putative CnfR-binding sites have been characterized in the promoter region of nifB [105, 106].

In *Anabaena*, expression of cnfR (patB) is induced in response to nitrogen deprivation [61, 103], exclusively in heterocysts [107]. Based on the presence of consensus NtcA-binding sites upstream of cnfR, it can be suggested that this gene is subjected to NtcA-dependent expression in both heterocyst-forming and N₂-fixing, non-heterocystous cyanobacteria [38, 105]. *Anabaena variabilis* harbors two nif gene clusters, nif1 that is expressed in the heterocysts and nif2 that is expressed in the vegetative cells under micro-oxic conditions (see e.g. [106]), and expression of the nif2 cluster regulatory gene cnfR2 has been shown to require NtcA [104]. Therefore, expression of cnfR2 is activated when combined nitrogen is scarce, but the Fe-S clusters may be formed only when the oxygen tension is low resulting in a protein that can promote expression of the nifB operon only under micro-oxic conditions [104, 105].

### 4.3.2. Expression of petH

The petH gene encodes ferredoxin-NADP⁺ oxidoreductase that is expressed at much higher levels in heterocysts than in vegetative cells [108], consistent with a very relevant role of this enzyme in the distribution of electrons in the heterocyst [19]. In heterocysts, petH is expressed from an NtcA-dependent promoter to which NtcA directly binds [109]. No good NtcA-binding site is however found in this promoter, suggesting the involvement of an NtcA coactivator such as PipX.

### 4.3.3. Expression of glnA

Glutamine synthetase, which incorporates into glutamate the ammonium produced in the N₂-fixation reaction to synthesize glutamine, is present at higher levels in the heterocysts than in vegetative cells [110]. The glnA gene encoding glutamine synthetase is expressed at higher
levels in the absence than in the presence of combined nitrogen, being transcribed from a
complex promoter region [111]. The promoter most proximal to the gene is a Class II NtcA-
dependent promoter that is used during growth with nitrate and under deprivation of
combined N [34, 35]. This is the *glnA* promoter that is used in the heterocysts, and its
dependence on NtcA has been corroborated by mutation of its NtcA-binding site [112].

4.3.4. Cyanophycin metabolism genes

As mentioned earlier, the dynamic N reservoir cyanophycin is accumulated conspicuously at
the heterocyst poles. In *Anabaena*, the genes encoding the main cyanophycin synthetase
(*all3879, cphA1*) and cyanophycinase (*all3880, cphB1*) are clustered together. These genes are
expressed as an operon (*cphB1-cphA1*) from a complex promoter region localized upstream of
cphB1, and cphA1 is also transcribed monocistronically from a complex promoter present in
the intergenic region. In these regions, NtcA-dependent promoters comprising high- and low-
affinity NtcA-binding sites are found which are active in heterocysts [113]. The gene *all3922,
encoding isoaspartyl dipeptidase that hydrolyzes the dipeptide β-aspartyl arginine, is induced
under nitrogen deprivation [61] and is expressed preferentially in vegetative cells [28],
consistent with the transfer of the dipeptide from the heterocysts to the vegetative cells.

4.3.5. Expression of RuBisCO

As mentioned earlier, the RuBisCO enzyme is not active in the heterocysts thus avoiding the
investment of energy and reductant in the fixation of CO$_2$, a process confined to the vegetative
cells. Lack of RuBisCO in the heterocysts results from repression of gene expression [114], and
binding of NtcA to a DNA site overlapping the translation start of the *rbcLXS* operon encoding
RuBisCO appears to represent a mechanism of repression [38, 115]. This mechanism of cell-
specific repression of RuBisCO stresses the relevant role of NtcA in regulation of gene
expression, not only by induction but also by repression, in the mature heterocyst.

5. Heterocyst patterning

Heterocyst distribution is not random in the cyanobacterial filaments, but instead it follows a
relatively fixed pattern. In most strains of the order Nostocales, heterocysts are separated by
to 15 vegetative cells along the filament. Importantly, this pattern is established early
upon N step-down, implying that products of N$_2$ fixation are not required for generation of at
least the initial pattern.
5.1. Coupled fluctuations of HetR expression

Several theoretical models have been put forward to try a mathematical description of the generation of the heterocyst distribution pattern, and most of those models are based on the operation of an activator and an inhibitor of differentiation with different intercellular diffusivities (revised in [116]). The HetR transcription factor has been considered as the activator in these models. As mentioned above, basal expression of \textit{hetR} takes place in the filament even in the presence of ammonium, and expression fluctuations are coupled in 2-4 nearby cells influenced by positive autoregulation ([117]; J. Stavans, personal communication). Moreover, this pattern persists during the initial stages of development upon N step-down [117], which could explain the classical observation that differentiation is first observed in small clusters of cells that are later resolved to single heterocysts [8]. Thus, coupled fluctuations of \textit{hetR} expression after N step-down would build-up upon the basic pattern already latent under N-replete conditions. At intermediate stages of differentiation, an increase in autocorrelation functions of \textit{hetR} expression is observed [117], which would result from increased \textit{hetR} autoregulation in response to \textit{ntcA} (and \textit{nrrA}) induction.

5.2. The PatS morphogen and the ABC exporter HetC

The \textit{patS} gene is induced early in specific cells, depending on HetR [46], and its overexpression abolishes heterocyst differentiation, whereas its inactivation leads to the formation of heterocysts in groups (the so called MCH, \textit{multiple contiguous heterocysts} phenotype). Hence, it was suggested that a \textit{patS} gene product effects a lateral inhibition of differentiation [118].

The primary product of the \textit{patS} gene is a 17-amino acid peptide that is processed to render a smaller active peptide including the C-terminal part of the primary gene product [118, 119], likely consisting of the C-terminal seven residues [119, 120]. Nonetheless, a peptide consisting of only the five C-terminal amino acids of PatS can also suppress heterocyst differentiation when added to \textit{Anabaena} cultures [118]. The active PatS peptide (which we will denote hereafter PatS\(^*\)) is transferred to the neighboring cells to establish a concentration gradient with highest levels in the cells adjacent to the producing cell, in which the concentration remains negligible, and decreasing thereon, thus behaving as a morphogen [119]. The N-terminal half of the \textit{patS} primary product is important for processing of the peptide and for conferring immunity against the inhibitory activity in the producing cell, which would involve peptide excretion [119].
The hetC gene encodes an ABC-type exporter that includes a putative peptidase domain required for heterocyst differentiation [121, 122]. hetC is expressed from a complex promoter region that includes a canonical NtcA-dependent promoter and a HetR- and NtcA-dependent promoter, similar to the promoter region of the devBAC operon [43, 89]. HetC localizes mostly to the heterocyst poles [122], and the possibility of its involvement in export of PatS* from the differentiating heterocysts has been considered [122, 123], as has also been considered the participation of the septal junction protein SepJ in the intercellular transfer of PatS* between vegetative cells [124]. Regarding the inhibitory activity of the morphogen, PatS* has been shown to interact with HetR [125, 126], to influence the turnover of HetR [127], and to inhibit HetR tetramer accumulation [53] and HetR binding to DNA [46]. All these observations are fully consistent with the reported isolation of a suppressor mutant of patS overexpression, which bore a mutation in the hetR gene [128]. Thus, at least one component of the inhibition of heterocyst differentiation by PatS* is the inhibition of HetR activity. Figure 4 presents a hypothetical model in which (i) HetC participates in processing of PatS and transfer of PatS* from the (pro)heterocyst to its neighboring vegetative cell, and (ii) SepJ participates in spreading of PatS* between vegetative cells along the filament. How PatS* is prevented from moving back to the (pro)heterocyst through SepJ-related conduits is unknown, but may be related to the distinct architecture of the septal junctions in vegetative cell-heterocyst septa described above.

Besides being dependent on hetR auto-regulation, the spatial correlation of hetR expression fluctuations along the filament, both in the presence of ammonium and early after N step-down, is influenced by expression of patS and intercellular communication dependent on SepJ [117]. This implies the existence of a basal expression of patS in the presence of ammonium, consistent with the reported detection of patS transcription in the presence of combined nitrogen [8]. Thus, fluctuations in the expression of hetR encoding a positive regulator of differentiation, HetR positive autoregulation, lateral inhibition by PatS (whose synthesis depends on HetR), and cell-to-cell communication are fundamental components of the developmental process.

5.3. Other proteins influencing pattern formation

Finally, other protein factors –including HetL, HetN, PatA, PatL and PatN– have been described to influence the spatial pattern of heterocyst distribution. The hetN gene influences the pattern of heterocysts especially at advanced stages of differentiation [129, 130]. This effect likely involves an internal hexapeptide sequence identical to the C-terminal sequence of PatS [130, 131]. Indeed, at late, but not at early times after N step-down, expression of hetN has
been shown to influence the coupling of \textit{hetR} expression fluctuations through the filament [117]. At times when heterocyst differentiation is already completed, the FraC/FraD proteins involved in intercellular molecular exchange impact \textit{hetR} expression correlations in neighboring cells in a manner similar to SepJ [117]. This effect of FraC/FraD could involve transfer of a HetN-derived product or other regulators or N$_2$ fixation-derived metabolites.

Whereas inactivation of \textit{patN} results in an increased heterocyst frequency [132, 133], inactivation of \textit{patA} [134] or \textit{patL} [135] results in the preferential production of terminal heterocysts in the filaments (implying a decreased heterocyst frequency). On the other hand, overexpression of \textit{hetL} increases heterocyst frequency [136]. PatL and HetL are pentapeptide repeat proteins of unknown function, PatA and PatL have been shown to interact with each other [135], and PatN has been shown to influence the expression of \textit{patA} [132]. Significantly, the \textit{patA} gene is transcribed in differentiating heterocysts from two NtcA-activated promoters [137]. PatA is a response regulator that lacks a DNA binding domain and counteracts PatS and HetN inhibition of heterocyst differentiation [138].

6. Concluding remarks

Heterocyst differentiation from a vegetative cell involves changes in expression of several hundred genes [139], including repression and induction (thoroughly reviewed by Xu \textit{et al.} [140]). Several transcription factors effecting this regulation have been identified mainly in the model organism \textit{Anabaena} (Table 1). NtcA and HetR are absolutely needed for heterocyst differentiation, whereas other transcription factors participate in specific steps of the differentiation process. NtcA and HetR are however not equivalent. Whereas NtcA is a global transcription factor that activates numerous gene promoters at different steps of differentiation, and represses some promoters as well, HetR has been shown to bind to few DNA sequences possibly activating transcription. Additionally, the strict requirement for HetR may rely on (i) its effect on localized induction of \textit{ntcA} in the developing heterocyst and (ii) a role as a coactivator of NtcA at specific promoters. An additional coactivator of NtcA is PipX, which being expressed late in heterocyst differentiation supports NtcA-dependent activation of late genes, including some that are expressed in mature heterocysts. In the mature heterocyst, the \textit{rbcL} operon promoter is directly repressed by NtcA, and a highly-expressed gene such as \textit{glnA} is transcribed from a canonical Class II NtcA-dependent promoter, raising the possibility that expression in the mature heterocyst no longer requires differentiation-specific transcription factors. This is consistent with an inactivation of HetR by phosphorylation.
as discussed above. Finally, in the mature heterocyst, expression of the *nif* genes requires a specific transcription factor (CnfR) that may link expression to the micro-oxic conditions prevailing in the differentiated cell.

**Acknowledgements**

Work in the authors’ laboratory is currently funded by grants number BFU2014-56757-P and BFU2017-88202-P (E.F.) and BFU2016-77097-P (AH) from Agencia Estatal de Investigación, Spain, co-financed by the European Regional Development Fund.
References


775 [57] Olmedo-Verd E, Muro-Pastor AM, Flores E, Herrero A, Localized induction of the ntcA regulatory
777 [58] Rajagopalan R, Callahan SM, Temporal and spatial regulation of the four transcription start sites of
779 [59] Muro-Pastor AM, Olmedo-Verd E, Flores E, All4312, an NtcA-regulated two-component response
781 [60] Ehira S, Ohmori M, NrrA, directly regulates expression of hetR during heterocyst differentiation in
783 [61] Flaherty BL, Van Nieuwerburgh F, Head SR, Golden JW, Directional RNA deep sequencing sheds new
784 light on the transcriptional response of Anabaena sp. strain PCC 7120 to combined-nitrogen
787 cyanobacterial PipX factor in the heterocysts of Anabaena sp. strain PCC 7120, J Bacteriol 193
788 (2011) 1172-82.
789 [63] Mitschke J, Vioque A, Haas F, Hess WR, Muro-Pastor AM, Dynamics of transcriptional start site
790 selection during nitrogen stress-induced cell differentiation in Anabaena sp. PCC7120, Proc Natl
792 [64] Llacer JL, Espinosa J, Castells MA, Contreras A, Forchhammer K, Rubio V, Structural basis for the
793 regulation of NtcA-dependent transcription by proteins PipX and PII, Proc Natl Acad Sci USA 107
796 and characterization of the cox3 promoter region in the heterocyst-forming cyanobacterium
798 [66] Brahamsha B, Haselkorn R, Isolation and characterization of the gene encoding the principal sigma
799 factor of the vegetative cell RNA polymerase from the cyanobacterium Anabaena sp. strain PCC
801 [67] Muro-Pastor AM, Brenes-Álvarez M, Vioque A, A combinatorial strategy of alternative promoter use
802 during differentiation of a heterocystous cyanobacterium, Environ Microbiol Rep 9 (2017) 449-
803 458.
804 [68] Imamura S, Asayama M, Sigma factors for cyanobacterial transcription, Gene Regul Syst Bio 3
806 [69] Khudyakov IY, Golden JW, Identification and inactivation of three group 2 sigma factor genes in
808 [70] Aldea MR, Mella-Herrera RA, Golden JW, Sigma factor genes sigC, sigE, and sigG are upregulated in
810 [71] Ehira S, Ohmori M, NrrA, a nitrogen-regulated response regulator protein, controls glycogen
811 catabolism in the nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120, J Biol Chem 286
814 normal expression of heterocyst-specific genes in Anabaena sp. strain PCC 7120, J Bacteriol 193
816 [73] Schrautemeier B, Neveling U, Schmitz S, Distinct and differently regulated Mo-dependent nitrogen-
817 fixing systems evolved for heterocysts and vegetative cells of Anabaena variabilis ATCC 29413:
818 characterization of the fdxH1/2 gene regions as part of the nif1/2 gene clusters, Mol Microbiol 18
820 [74] Saha SK, Golden JW, Overexpression of pknE blocks heterocyst development in Anabaena sp. strain


Figure 2. Mechanism of activation of AlgU (RpoE)

No stress

<table>
<thead>
<tr>
<th>AlgW (DegS)</th>
<th>AlgU (σE)</th>
<th>ClpX/P</th>
</tr>
</thead>
<tbody>
<tr>
<td>MucB (RseB)</td>
<td>MucA (RseA)</td>
<td></td>
</tr>
</tbody>
</table>

Cell wall stress

1. Misfolded porins
2. CupB5
3. MucP (RseP)

<table>
<thead>
<tr>
<th>AlgU</th>
<th>T4Pili</th>
<th>LPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MucD</td>
<td>MucE</td>
<td>MucP</td>
</tr>
</tbody>
</table>

OM

Periplasm

IM

Cytoplasm

Regulon

Proteolytic domain

PDZ domain
Figure 3: ECF sigma factors involved in regulators expression
Figure 4: Regulators involved in ECF sigma factors expression
Figure 5. FpvA cell surface signaling

Fe^{3+}

Pyoverdine

OM

Periplasm

CM

Cytoplasm

Pvd genes

toxR
prpl
pyoS3A
exoT
exsA
hemH

fpvA
pvdA
pvdD
pvdI
pvdJ
fpvGHJK
hasR
prf1
Figure 6: Pyoverdine cluster
Regulatory interaction

Connectivity

ECF sigma factors displayed in different colors

Anti-ECF sigma factors

Sigma factors

Two-component regulatory systems

Transcriptional regulators

Non-coding RNA gene (sRNA)