PNAS RESEARCH ARTICLE | MICROBIOLOGY OPEN ACCESS

Coordination of N_2-fixing cell specialization and patterning in filamentous cyanobacteria by a uniquely structured σ factor

Xi Hu[a,b,1], Xujiao Liu[c,d,1], Yali Wang[a,1], Ke Liu[a], Hong Gao[b], Geqian Zhu[b,d], Shuai Wang[b], Xiantao Fang[b], Hong Sun[c,d], Fang Zhou[b], Shuiling Ji[a], Congzhao Zhou[e], Yu Zhang[c,2], and Xudong Xu[a,b,2]

Edited by Susan S. Golden, University of California San Diego, La Jolla, CA; received March 21, 2026; accepted June 7, 2026

Cell differentiation and Turing-like patterning are tightly associated in a group of filamentous cyanobacteria that differentiate specialized N_2-fixing cells, called heterocysts. Based on systematic genetic analyses, in particular genome-wide identification of recognized promoters and assays with a reconstituted *Anabaena* transcription system in *Escherichia coli*, we established HetZ as the central activator of the gene regulatory network of heterocyst differentiation. Biochemical and cryo-EM analyses further established HetZ as a σ factor (σ^{HetZ}). Unique domain insertions in σ^{HetZ} are involved in promoter DNA recognition-unwinding and interaction with the inhibitor PatU3. σ^{HetZ}-PatU3 and the master regulator-diffusible inhibitor constitute the minimal core regulatory circuit (CRC) for cell fate determination and patterning. σ^{HetZ} activates not only genes of the CRC but also downstream regulator/effector genes involved in morphological and functional development. The gene regulatory network and the structure–function relationship of σ^{HetZ} depict how cell differentiation and patterning are coordinated in this group of multicellular cyanobacteria.

cell differentiation | multicellular patterning | core regulatory circuit | gene regulatory network | unique σ factor

Significance

A uniquely structured σ factor is established as the central activator of N_2-fixing cell specialization in a group of globally important photosynthetic diazotrophs, providing the hitherto most extensive depiction of the gene regulatory network of cyanobacterial heterocyst differentiation. The findings also extend our understanding of the role of σ factors to Turing-like multicellular patterning.

Multicellular bacteria evolve at a higher organization level than unicellular ones. For problems typically encountered in bacteria, such as inactivation of nitrogenase by O_2, multicellular bacteria could evolve a solution through cell differentiation and patterning. This is the case for heterocyst-forming cyanobacteria, including many free-living species and some symbionts of diatoms or plants. Heterocysts are specialized N_2-fixing cells differentiated from vegetative cells that otherwise perform oxygenic photosynthesis to provide micro-oxic intracellular environment for nitrogenase (1–3). Heterocyst differentiation is tightly associated with Turing-like patterning along multicellular filaments (1, 4). Heterocyst-forming cyanobacteria contribute significantly to nitrogen fixation in oceanic, limnetic, and terrestrial ecosystems (5–7), and may be used to generate vertically transferred endosymbionts in plants via synthetic biological manipulations. Hitherto, the core regulatory circuit (CRC) for heterocyst differentiation/patterning has not been defined, and accordingly, the gene regulatory network of heterocyst differentiation (het-GRN) remains to be integrated into one picture.

During heterocyst differentiation, a thickened envelope consisting of an outer polysaccharide layer and an inner glycolipid layer is formed to limit the entry of oxygen (8, 9), whereas photosynthetic oxygen evolution is greatly diminished or lost (10), and oxygen consumption is substantially intensified (11, 12). Heterocyst differentiation has been extensively analyzed in the model strain *Anabaena* sp. PCC 7120 (hereafter *Anabaena* 7120). In response to nitrogen stepdown, single heterocysts are produced at a semiregular pattern in uniseriate filaments of *Anabaena* 7120, and the differentiation process can be divided into three stages. At the first stage, the induction of heterocyst differentiation is mediated by the accumulation of 2-oxoglutarate (2-OG), which binds to NtcA, enhancing the DNA-binding activity of this CRP-family global nitrogen regulator (13, 14). In turn, NtcA indirectly upregulates the expression of *hetR*, the master regulator gene of heterocyst differentiation (15, 16). This stage is missing in species that differentiate heterocysts irrespective of nitrogen compound availability, such as *Cronbergia* (17). At the second stage, cell differentiation and Turing-like patterning are initiated, depending on the interaction between HetR and its inhibitor(s), diffusible RG(S/T)GR-containing peptides derived from the 17-aa PatS (18, 19) and the 77-aa PatX (20). In *Anabaena* 7120, PatS (RGSGR) is the major contributor of such peptides for de novo pattern formation (21, 22). In addition to HetR-RGSGR, there is another activator–inhibitor pair, HetZ-PatU3,

Author affiliations: [a]Key Laboratory of Pesticide and Chemical Biology of Ministry of Education, Hubei Key Laboratory of Genetic Regulation and Integrative Biology, School of Life Sciences, Central China Normal University, Wuhan 430079, China; [b]Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, China; [c]Key Laboratory of Synthetic Biology, Key Laboratory of Plant Design, Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China; [d]University of Chinese Academy of Sciences, Beijing 100049, China; and [e]School of Life Sciences and Biomedical Sciences and Health Laboratory of Anhui Province, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230026, China

Author contributions: X.H., X.L., Y.W., Y.Z., and X.X. designed research; X.H., X.L., Y.W., K.L., H.G., G.Z., S.W., X.F., H.S., F.Z., and S.J. performed research; X.H., X.L., Y.W., K.L., C.Z., Y.Z., and X.X. analyzed data; and X.H., X.L., Y.W., C.Z., Y.Z., and X.X. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]X.H., X.L., and Y.W. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: yzhang@cemps.ac.cn or xuxudong@ccnu.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2610149123/-/DCSupplemental.

Published July 1, 2026.

involved in the pattern formation. PatU5 and PatU3 in *Anabaena* 7120 correspond to the 5′ and 3′ portions of PatU in most other filamentous cyanobacteria. HetR directly regulates the expression of *hetZ* (23); conversely, the heterocyst-specific expression of *hetR*, *patS,* and *patX* appears to depend on HetZ (24). However, no evidence has been presented to show the direct regulation of these genes by HetZ. In contrast, the regulation by HetZ is either thought to be indirect (25) or overlooked, with HetZ designated as a factor for "commitment" of heterocyst differentiation (26, 27). At the third stage, heterocysts are morphologically and functionally developed. Genes for polysaccharide (28, 29) and glycolipid (30–32) layers, pattern maintenance (33), and genes for the activities of oxygen consumption (11, 12) and N_2 fixation (34), mostly organized in gene clusters, are upregulated sequentially in developing heterocysts. At this stage, NtcA regulates glycolipid layer formation, oxygen consumption, and nitrogen fixation directly (35) or indirectly (12, 36).

Most often, the mutual regulation between NtcA and HetR, and the regulation between HetR and PatS, are depicted as the central machinery for initiating heterocyst differentiation/patterning (26, 27). However, neither *patS* nor *ntcA* is directly regulated by the master regulator HetR. There must be a factor that directly mediates and coordinates these regulatory relationships. In this study, we successfully reconstructed the framework of het-GRN based on multiple lines of evidence, including Chromatin Immunoprecipitation Sequencing (ChIP-seq), RNA sequencing (RNA-seq), the use of a reconstituted heterocyst transcription system in *Escherichia coli* and genetic analyses in *Anabaena*. In the het-GRN, HetZ activates the expression of *hetR*, *patS*, *patU5-patU3*, *ntcA,* and downstream genes. Cryo-electron microscopy (EM) structure analysis revealed that HetZ is an ECF-type sigma factor (σ^{HetZ}) with unique domain insertions involved in promoter DNA recognition and unwinding. Using a synthetic biological strategy, we showed that the two activator–inhibitor pairs, HetR-PatS and σ^{HetZ}-PatU3, constitute the minimal CRC that determines the cell fate and de novo patterning.

Results

Central Role of HetZ (σ^{HetZ}) in Heterocyst Differentiation. To determine how HetZ is involved in heterocyst differentiation, we performed ChIP-seq analyses with rabbit anti-HetZ IgG antibody in *Anabaena* 7120 and the *hetZ*4-201 mutant [with no. 2 to 67 codons deleted (24)] to identify the specific genomic targets of HetZ. ChIP-seq peaks significantly decreased in the mutant relative to the wild type (*SI Appendix*, Fig. S1*A*, based on Log_2 *hetZ*4-201/WT; Datasets S1 and S2) were considered to be specific to HetZ recognition. Some genes downstream of the differential peaks are involved in 2-OG sensing, cell fate determination, and heterocyst morphogenesis. These genes include *ntcA* (13, 14), *hetR* (15, 16), *patS* (18, 19), *patU5-patU3* (24, 37), *sigC* (σ^{C}), and *alr2825* through *alr2841* (the *hep* gene cluster for polysaccharide layer formation) (28) (Fig. 1*A*). Other genes related to heterocyst differentiation, such as *patX* (77-aa peptide, producing RGTGR inhibitor of HetR) (20) and *hetC* (peptidase-based ABC exporter) (38), are indicated in the *SI Appendix*, Fig. S1*A*.

TCCGGA was initially identified as the *cis*-element, called DIF1 motif, of genes downregulated in a *hetR* mutant compared to the wild type (39), but the activation of DIF1-containing promoters was shown to be more directly dependent on *hetZ* rather than *hetR* (24). In 546 differential peaks (Dataset S2), including those upstream of *hetR*, *patS*, *patU5-patU3*, *sigC,* and *hep* genes, we identified the consensus sequence of the DIF1-containing promoters (Fig. 1*C* and *SI Appendix*, Fig. S1*C*). The transcription start site of *ntcA* (40) is also located in the ChIP-seq peak shown in Fig. 1*A*, even though the −35 and −10 regions are less consistent with a typical DIF1-containing promoter (Fig. 1*C*). RNA-seq analyses indicate that *ntcA*, *patS*, *sigC,* and *hep* genes are downregulated in the *hetZ* mutant relative to the wild type (*SI Appendix*, Fig. S1*B*). However, the effects of HetZ on the expression of *hetR* and *patU5-patU3* are masked by the transcription from other promoters. Apparently, HetZ-recognition sites are proximal to the transcription start sites and composed of two separate sequence motifs spaced by 16-17 base pairs, indicative of HetZ as a σ factor. The Dali structure similarity search of the AlphaFold2-predicted HetZ structure suggests that it contains domains similar to bacterial ECF σ factors (41, 42). Thus, we tentatively name HetZ as σ^{HetZ}.

NtcA plays important roles at stages I (N deficiency sensing) and III (morphogenesis) of heterocyst differentiation, during which it indirectly upregulates *hetR* (13, 14) and depends on HetR for upregulation (43). The ChIP-seq result indicated the upregulation of *ntcA* is probably activated by σ^{HetZ} (Fig. 1*A*). We performed genetic analysis using green fluorescent protein gene (*gfp*) fused to the promoter of *ntcA*. P_{ntcA}-*gfp* showed heterocyst-specific expression, and the expression was abolished by mutations at the −35 region of P_{ntcA} or inactivation of *hetZ* (Fig. 1*B*). Furthermore, P_{ntcA}-*gfp* was upregulated in heterocysts without HetR, which were produced by expressing *hetZ* (σ^{HetZ}) from the copper-regulated promoter P_{petE} in the Δ*hetR* mutant (Fig. 1*B*). Apparently, σ^{HetZ} activates the expression of *ntcA* through a less typical σ^{HetZ}-recognized promoter (Fig. 1*C*). Actually, as shown with in vitro transcription assays described later, an active σ^{HetZ}-recognized promoter could have alternative bases at some positions of −35 and −10 regions. The σ^{HetZ}-recognized promoter of *ntcA* is consistent with the sequence required for active transcription except for the first base of −35 region (Fig. 1*C*). For this reason, an additional regulator may be required for the σ^{HetZ}-activated expression of *ntcA* in *Anabaena* 7120.

Now that σ^{HetZ} directly activates *hetR*, *patS* and *patX*, *patU5-patU3*, *ntcA*, *sigC,* and the *hep* gene cluster, then what is the role of the nonspecific DNA-binding protein HetP that functionally overlaps with HetZ (44)? In the Δ*hetP* mutant, σ^{HetZ} was detected at a higher level than in the wild type (*SI Appendix*, Fig. S1*D*), but all the above σ^{HetZ} ChIP-seq peaks disappeared or were greatly diminished (Fig. 1*A*). RNA-seq analysis was also performed to compare the transcriptional profiles in the Δ*hetP* mutant and the wild type, and the result was similar to that of *hetZ*4-201 vs. wild type (*SI Appendix*, Fig. S1*B*). It is likely that HetP increases the accessibility of target DNA sites to σ^{HetZ}, therefore enhances the transcription of most σ^{HetZ}-activated genes in *Anabaena*.

The Gene Regulatory Network of Heterocyst Differentiation (het-GRN) with σ^{HetZ} as the Central Activator. To systematically elucidate the regulatory network of heterocyst differentiation, we reconstituted the transcription system of *Anabaena* in *E. coli*. In such a system, the tested protein is the only potential regulatory factor from *Anabaena*, therefore can be interpreted as the direct cause for the activation of promoter-*gfp* fusion. The *E. coli* system comprises 2 or 3 plasmids. The three-plasmid system includes the plasmid pHB7508, in which *Anabaena* RNAP genes (except the ω subunit gene) are controlled by the arabinose-inducible promoter P_{araB}, the plasmid with *hetZ* expressed from P_{cat} on pACYC184, and the plasmid with promoter-*gfp* (Fig. 2 *A*–*I*). In *E. coli* DH10B containing the three plasmids, σ^{HetZ} activated the expression from P_{hetR}, P_{patS}, $P_{patU5\text{-}patU3}$, P_{sigC}, $P_{alr2830}$, $P_{alr2833}$, P_{hepA} (Fig. 3*A* and *SI Appendix*, Fig. S2); mutations of the −35 element

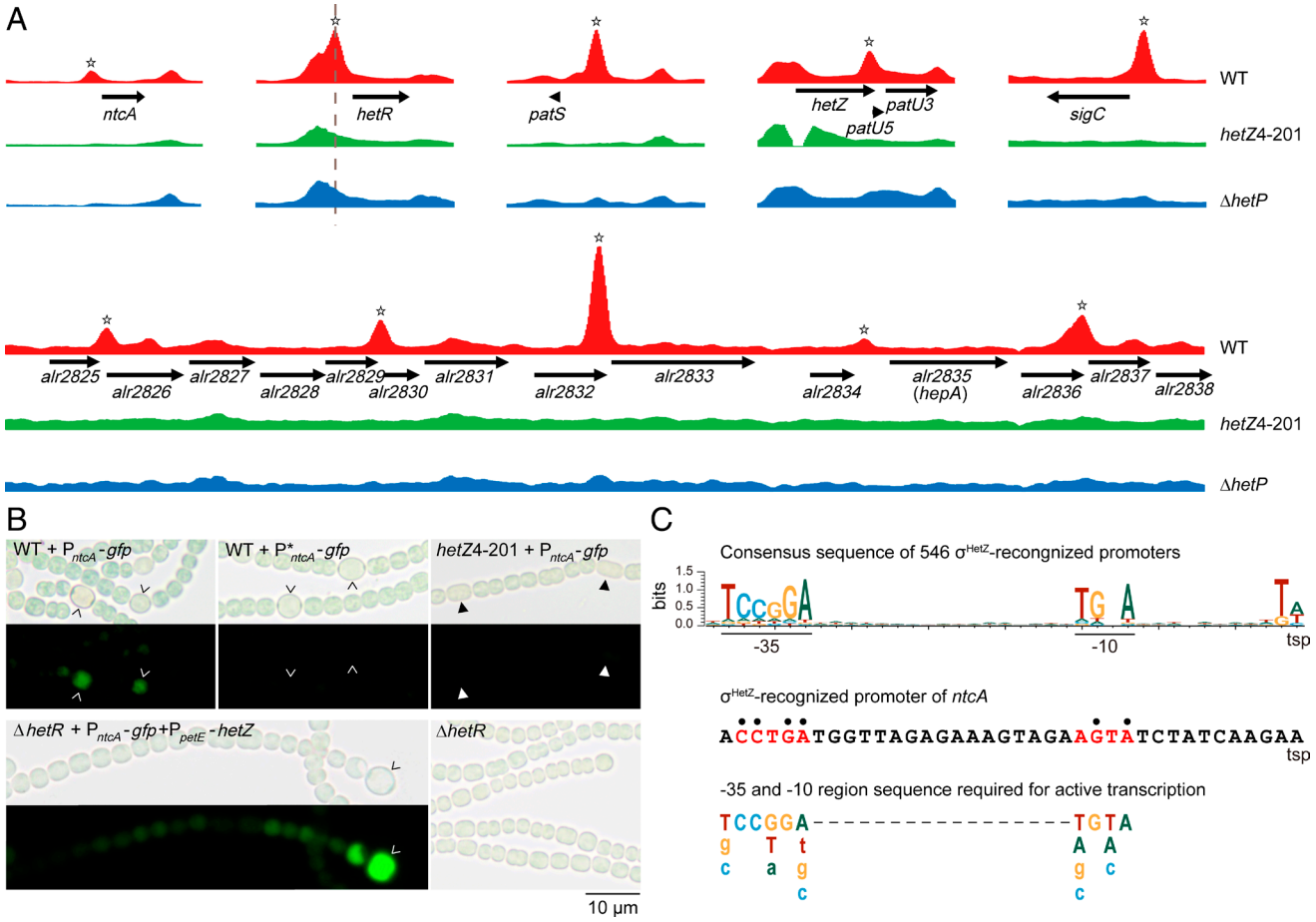

Fig. 1. Direct activation of genes involved in heterocyst differentiation and patterning by HetZ (σ^{HetZ}). (*A*) ChIP-seq peaks of HetZ (empty stars) upstream of genes involved in heterocyst differentiation or patterning in the wild type and, the *hetZ*4-201 mutant and the Δ*hetP* mutant at 8 h after nitrogen stepdown. Genome-wide peaks are listed in Dataset S1. HetZ is proposed to be σ^{HetZ} for reasons described in the text. (*B*) Light (*Upper*) and GFP fluorescence (*Lower*) photomicrographs of *Anabaena* 7120 strains at 24 h (derivatives of WT and Δ*hetR*) or 48 h (derivative of *hetZ*4-201) after nitrogen stepdown, showing σ^{HetZ}-dependent expression from P_{ntcA} in heterocysts. P^*_{ntcA}, P_{ntcA} with substitutions at the −35 region. WT, wild type. Empty arrowheads point to heterocysts, solid arrowheads indicate cells that initiated differentiation but ceased at an early stage. The leaky phenotype of *hetZ*4-201 is due to the presence of *hetZ′* in the genome as described later. (*C*) Comparison of the σ^{HetZ}-recognized promoter of *ntcA* with the consensus sequence of 546 σ^{HetZ}-recognized promoters and the −35 and −10 region sequence required for active transcription. In the promoter of *ntcA*, bases in bright red are consistent with the sequence for active transcription, those indicated with a dot coincide with the consensus sequence of 546 promoters. In the −35 and −10 region sequence for active in vitro transcription, uppercase letters stand for bases with >60% activities, lowercase letters for 20 to 60% activities. tsp, transcription start point.

("–" in Fig. 3*A*, P* in *SI Appendix*, Fig. S2), or use of *hetZ* mut-1 instead of *hetZ* (σ^{HetZ}-P_{hetR} 3P *iii* in *SI Appendix*, Fig. S2), greatly decreased the expression of *gfp* from these promoters. The efficacy of the three-plasmid system was confirmed with a two-plasmid system (Fig. 2 *A*, *III*, based on pHB8140), with which IPTG-induced expression of P_{lacZ}-*hetZ* directly activated the expression of P_{hetR}-*gfp* but not if the −35 element of P_{hetR} was mutated (Fig. 2 *A*, *III*; σ^{HetZ}-P_{hetR} 2P in *SI Appendix*, Fig. S2). These results provide direct evidence that σ^{HetZ} activates genes of the CRC, namely *hetR*, *patS*, *patU5-patU3* (*patU*), and genes involved in heterocyst morphogenesis, in particular *hep* genes and *sigC*. The less typical σ^{HetZ}-recognized promoter of *ntcA* was not activated in the *E. coli* system. Probably, an additional regulator is required for activation of this promoter of *ntcA*.

In addition to assays with the *E. coli* system, we also checked σ^{HetZ}-promoted expression from $P_{patU5\text{-}patU3}$, P_{sigC}, $P_{alr2830}$, $P_{alr2833}$, P_{hepA} in *Anabaena* 7120. *gfp* fused to these promoters was specifically expressed in heterocysts, and the expression was abolished or greatly decreased by mutations at the −35 region or inactivation of *hetZ* (*SI Appendix*, Fig. S3). The regulatory relationships revealed by the reconstituted system in *E. coli* (Fig. 3*A* and *SI Appendix*, Fig. S2) and the native system in *Anabaena* (Fig. 1*B* and *SI Appendix*, Fig. S3) are summarized in the het-GRN of Fig. 2*B*. Indeed, σ^{HetZ} serves as the central activator in the heterocyst differentiation network.

PatU3 has been shown to interact with σ^{HetZ}, counteracting the effect of σ^{HetZ} on gene expression in *Anabaena* 7120 (24); *patU5-patU3*, as genes corresponding to the 5′ and 3′ portions of *patU* in other filamentous cyanobacteria, are cotranscribed with *hetZ* (37). All these features indicate that PatU3 is likely to be an anti- σ factor for σ^{HetZ}. Consistently, σ^{HetZ}-activated expression in the *E. coli* system was inhibited by PatU3 (Fig. 3*A* and *SI Appendix*, Fig. S2).

In heterocyst-forming cyanobacteria, there is a second σ^{HetZ}, called $\sigma^{HetZ'}$, which is encoded by *hetZ′* (*alr0202*) (*SI Appendix*, Fig. S4*A*). This is why the *hetZ*4-201 mutant shows a "leaky" phenotype, with some cells initiating the differentiation under nitrogen-depleted conditions. Unlike the single mutant, a Δ*hetZ′*P_{petE}-*hetZ* double mutant shows no cell differentiation in copper-free (turning off P_{petE}) nitrogen-deficient medium (*SI Appendix*, Fig. S4*B*). In the *E. coli* system, $\sigma^{HetZ'}$ promoted the expression of *gfp* from P_{hetR}, P_{patS}, $P_{patU5\text{-}patU3}$, P_{sigC}, $P_{alr2830}$, and P_{hepA}, and the activity of $\sigma^{HetZ'}$ was counteracted by PatU3 (*SI Appendix*, Fig. S2). Thus, σ^{HetZ} and $\sigma^{HetZ'}$ are almost identical

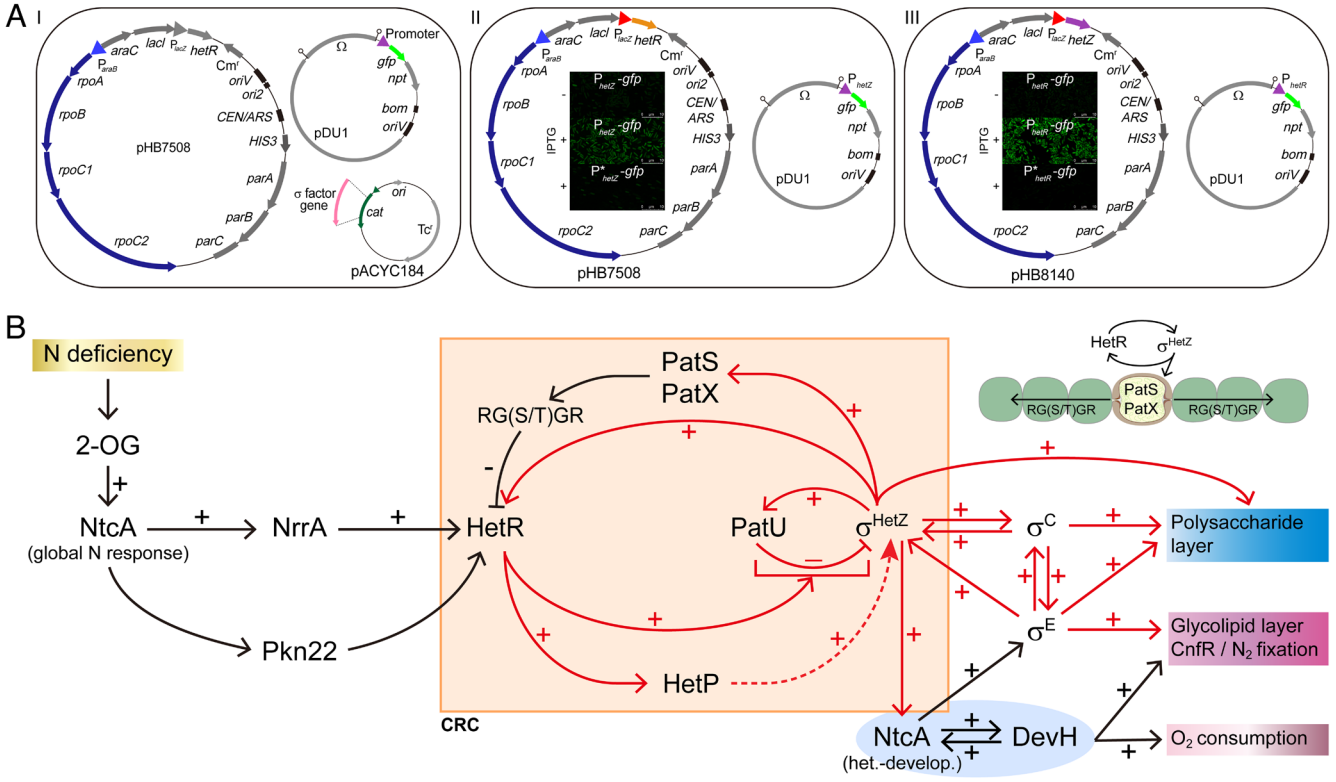

Fig. 2. Reconstruction of the het-GRN. (*A*) Heterocyst transcription system reconstituted in *E. coli* DH10B. (*I*) A three-plasmid system based on pHB7508, used to detect the activation of promoters by σ^{HetZ}. In the presence of arabinose and the absence of IPTG, genes encoding *Anabaena* RNAP core subunits were expressed on pHB7508, while *hetR* was turned off. One of the σ factors (σ^{HetZ}, $\sigma^{HetZ'}$, σ^{C} or σ^{E}) expressed from genes cloned in pACYC184 turned on the expression of promoter-*gfp* on the third plasmid; the empty vector pACYC184 was used as the negative control. (*II*) A two-plasmid system based on pHB7508, with GFP fluorescence photomicrographs of *E. coli* cells showing the regulation of P_{hetZ}-*gfp* by HetR. P^*_{hetZ}, P_{hetZ} with substitutions at the HetR-binding site. IPTG induced the expression of P_{lacZ}-*hetR*. (*III*) A two-plasmid system based on pHB8140, with GFP fluorescence photomicrographs of *E. coli* cells showing the activation of P_{hetR}-*gfp* by σ^{HetZ}. P^*_{hetR}, P_{hetR} with substitutions at the −35 region. IPTG induced the expression of P_{lacZ}-*hetZ*, producing σ^{HetZ}. (*B*) A schematic diagram showing the het-GRN, with σ^{HetZ} acting as the central activator. The het-GRN presented here is intended to summarize the results of this study, indicated in red, and regulation mediated by NtcA (14, 35, 45, 46), NrrA (47, 48), Pkn22 (49), and DevH (36, 50). PatU in filamentous cyanobacteria splits into PatU5 and PatU3 in *Anabaena* 7120. Lines with empty arrowhead stand for activation or induction, flat head for inhibition; the dashed line with solid arrowhead stands for enhancement. + and −, positive and negative regulation. 2-OG, 2-oxoglutarate. The cartoon at the upper right corner shows a heterocyst and neighboring vegetative cells, adding spatial information to the het-GRN. Arrowheaded lines across vegetative cells indicate the diffusion of RG(S/T)GR peptides along the filament.

in function. However, *hetZ'* is expressed at a much lower level (*alr0202* in *SI Appendix*, Fig. S1*B*) and in a different regulatory mode (51) compared to *hetZ*.

Two other σ factors, σ^C and σ^E (previously SigC and SigE), have been shown to be involved in gene regulation during heterocyst differentiation (52–54), but it is not clear which genes are their direct targets. Assays with the *E. coli* system indicated that σ^C activates the expression from P_{hetZ}, P_{sigE}, and $P_{alr2830}$, whereas σ^E activates the expression from P_{hetZ}, P_{sigC}, $P_{alr2830}$, P_{hglD}, and P_{cnfR} (Fig. 3*B* and *SI Appendix*, Fig. S2). *hglD* is involved in the synthesis of heterocyst glycolipids; *cnfR*, also called *patB*, encodes the regulator of *nif* genes for synthesis of nitrogenase. Thus, σ^{HetZ}, σ^C and σ^E form a "3- σ" mutual regulation circuit; through σ^C and σ^E, σ^{HetZ} controls *nif* genes for N_2 fixation and the *hglD* gene for glycolipid synthesis (Fig. 2*B*).

We also tested the regulation of *hetZ* and other genes by HetR, the master regulator. In *E. coli* harboring pHB7508 and a plasmid with promoter-*gfp* (Fig. 2 *A*, *II*), IPTG induced the expression of HetR, then HetR activated the expression of *gfp* from P_{hetZ} and P_{hetP}, presumably relying on an endogenous *E. coli* σ factor; substitutions at the HetR-binding site abolished the expression from P_{hetZ} and P_{hetP}(*SI Appendix*, Fig. S2). No expression was observed from P_{hetR}, P_{patS}, and P_{hepA} (*SI Appendix*, Fig. S2). Apparently, the synthetic transcription system in *E. coli* is not only useful in assays of σ factors but also in assays of regulators like HetR.

CRC for Cell Fate Determination. In the het-GRN, HetR, σ^{HetZ}-PatU3, HetP, PatS/PatX constitute the CRC for heterocyst differentiation and de novo pattern formation. To substantiate the role of the CRC in initiation of cell differentiation and patterning, we first generated a septuple mutant of *Anabaena* 7120, in which *hetR*, *hetZ*, *patU5-patU3*, *hetP*, *patS*, *patA*, and most of the *hgl* gene clusters were deleted. *patA* is a gene required for heterocyst formation at intercalary positions in *Anabaena* 7120 (55), activated by HetR in vegetative cells (51). Using this septuple mutant strain as a nonheterocystous "chassis," we tested the efficacy of reconstituted CRC modules.

First, we tested the mini-CRC with *hetR* and *hetZ-patU5-patU3* on an RSF1010-based plasmid and *patX* on the chromosome. These genes encode two activator–inhibitor pairs, HetR-PatX and σ^{HetZ}-PatU3, that constitute the core regulatory loops. The expression of *hetR* is reported with *ofp* (orange fluorescent protein gene), while *hetZ-patU5-patU3* is reported with *gfp* (Fig. 4 *A*–*I*). HetP is not required for pattern formation (56), therefore not included in the mini-CRC. Under nitrogen-depleted conditions, the septuple mutant with the mini-CRC produced multiple contiguous enlarged cells in which *hetR* and *hetZ-patU5-patU3* were upregulated (Fig. 4 *A*, *II*). These cells were differentiating heterocysts without the glycolipid layer. Second, we added *patS* to the mini-CRC (Fig. 4 *B*, *I*). The presence of *patS* inhibited the enlargement of differentiating cells and reduced the number of

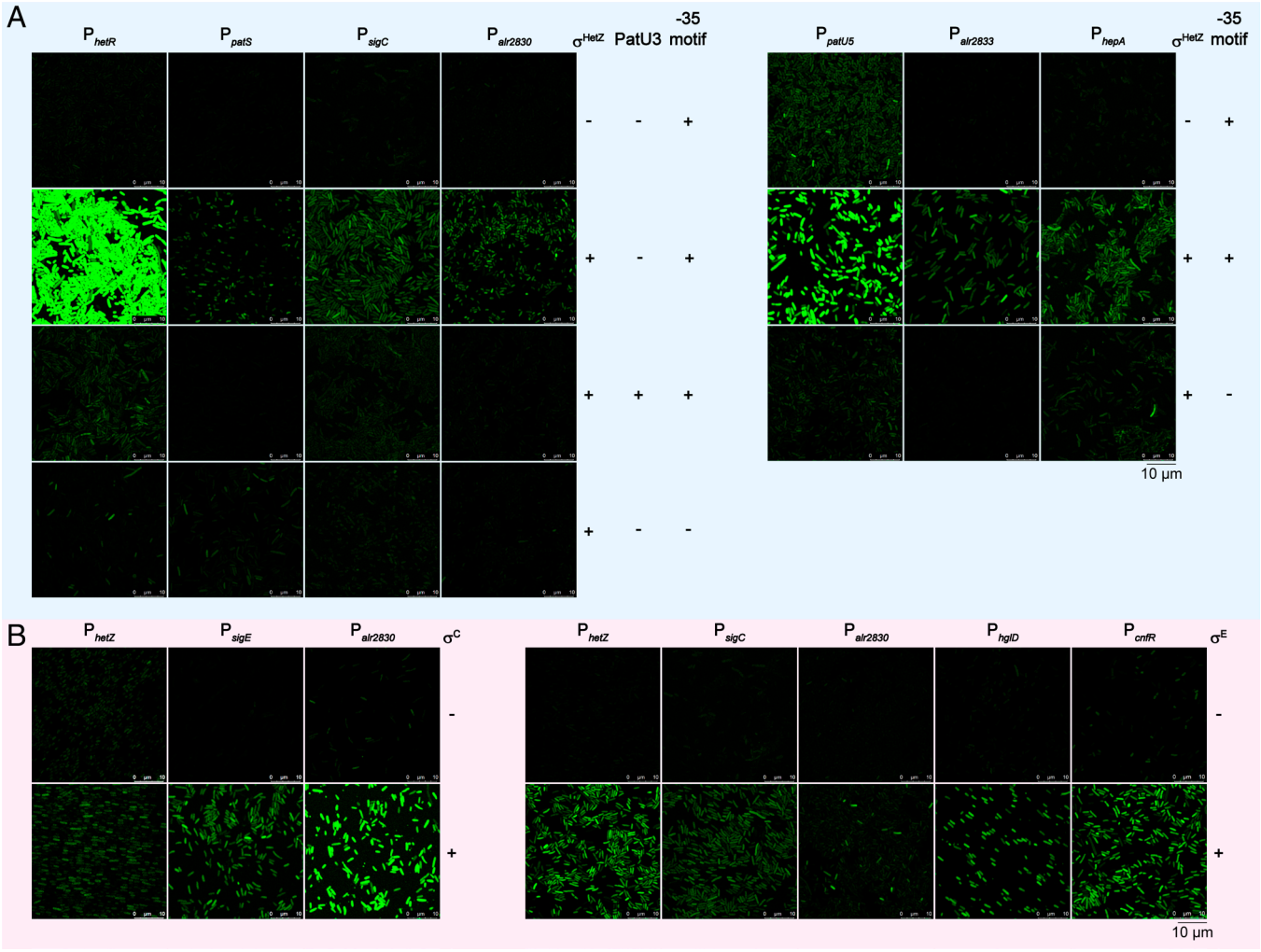

Fig. 3. GFP fluorescence photomicrographs of *E. coli* cells with the *Anabaena* transcription system, showing regulatory relationships in the het-GRN. In the *E. coli* assay system with pHB7508, P_{cat}-*hetZ* (encoding σ^{HetZ}), P_{cat}-*hetZ-patU3*, P_{cat}-*sigC* (encoding σ^{C}), or P_{cat}-*sigE* (encoding σ^{E}), was carried on the second plasmid, and promoter-*gfp* on the third plasmid. −35 motif, TCCGGA allowing 1-2 mismatches; + and −, with or without a protein (the second plasmid with or without the encoding gene) or a *cis*-element (the wild-type or mutated regulatory sequence). (*A*) Activation of promoters by σ^{HetZ}. (*B*) Activation of promoters by σ^{C} or σ^{E}.

contiguous differentiating cells (Fig. 4 *B*, *II*). Substitutions at the HetR-binding site of *hetZ*, therefore loss of the promoter activity, abolished the initiation of cell differentiation in both cases (Fig. 4 *A*, *III* and Fig. 4 *B*, *III*). These results clearly indicated that the mini-CRC is sufficient for initiating cell differentiation and pattern formation. It is noteworthy that cell differentiation was initiated at intercalary positions without *patA* in such reconstituted strains.

The Structure of σ^{HetZ}-RNAP-Promoter Open Complex. To definitely establish HetZ as a σ factor that coordinates heterocyst differentiation and patterning, we investigated the structural basis for HetZ to activate the expression from a typical promoter. First, we tested whether the tentative σ^{HetZ} could assemble with the RNAP core enzyme to form an RNAP holoenzyme and performed in vitro transcription from the promoter of *patS*. In size-exclusion chromatography, σ^{HetZ} was eluted within the same peak of the RNAP core enzyme, indicative of RNAP- σ^{HetZ} holoenzyme formation (*SI Appendix*, Fig. S5*A*). In the in vitro transcription system, RNAP- σ^{HetZ} produced the full-length transcript from P_{patS} but not from T5-N25, a representative σ^{A}-type phage promoter (*SI Appendix*, Fig. S5*B*). Sequence alignment of σ^{HetZ} with other σ factors suggests that σ^{HetZ} belongs to a unique subgroup of ECF σ factors containing three domain insertions, namely σ^{HetZ}_{i1}, σ^{HetZ}_{i2}, and σ^{HetZ}_{i3} (Fig. 5*D* and *SI Appendix*, Fig. S6*A*). A stretch of sequence including the entire σ^{HetZ}_{i2} and $\sigma^{HetZ}_{3.2}$ domains shows similarity to the nucleotide-interacting region of the single-stranded DNA–binding protein TcaR in staphylococci (*SI Appendix*, Fig. S6*B*) (57–59), indicating involvement of some recombination events in the origination of σ^{HetZ}.

We then determined the cryo-EM structure of the σ^{HetZ}-RNAP open complex (σ^{HetZ}-RPo) (*SI Appendix*, Fig. S5 *C*–*F*). The complex was reconstituted using *Anabaena* 7120 RNAP core enzyme, σ^{HetZ}, a partially pre-unwound *patS* promoter DNA, and a 5-nt RNA primer (Fig. 5*A*). The σ^{HetZ}-RPo structure was determined using a single-particle cryo-EM method at a resolution of 2.88 Å. The cryo-EM map exhibited clear signals for six subunits (α^{I}, α^{II}, β, β′1, β′2, ω) of the RNAP core enzyme and the six domains of σ^{HetZ} (σ^{HetZ}_{i1}, σ^{HetZ}_{2}, σ^{HetZ}_{i2}, $\sigma^{HetZ}_{3.2}$, σ^{HetZ}_{4}, and σ^{HetZ}_{i3}) (Fig. 5 *B*–*D*). The cryo-EM map also revealed clear densities of the upstream (−40 to −13) and downstream (+3 to +15) double-stranded DNA (dsDNA), the complete single-stranded nontemplate DNA of the transcription bubble, a portion of the single-stranded template DNA (−6 to +2) of the transcription bubble, and the 5-nt RNA primer base-paired with template DNA (−4 to +1) (Fig. 5*E*).

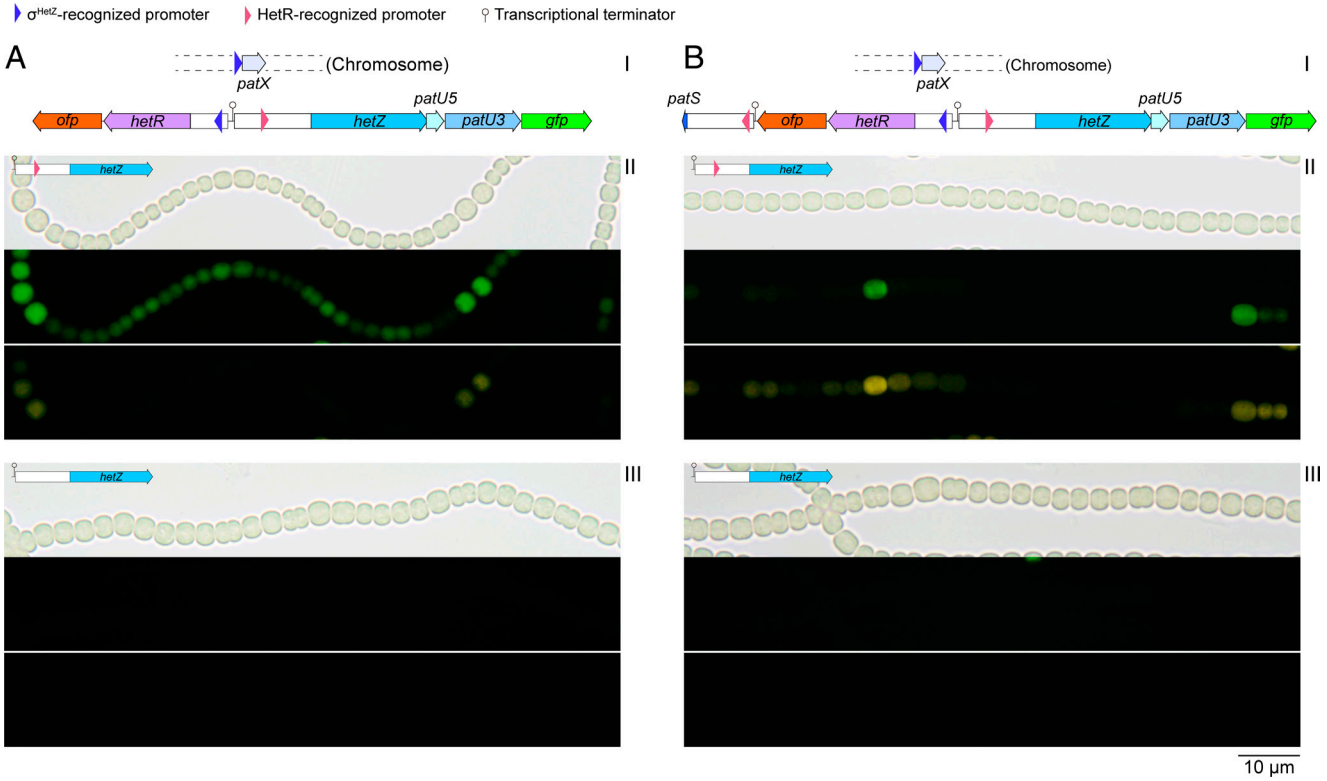

Fig. 4. Cell differentiation based on the mini-CRC. (*A* and *B*) The *gfp-/ofp*-labeled mini-CRC without or with *patS*. (*I*) Schematic diagram showing the reconstituted mini-CRC. *patU5-patU3* corresponds to *patU* in most other filamentous cyanobacteria. (*II* and *III*) Light (*Upper*), GFP (*Middle*), and OFP (*Lower*) fluorescence photomicrographs showing cell differentiation in the septuple mutant of *Anabaena* 7120 harboring the mini-CRC, with or without the HetR-recognized promoter of *hetZ*, at 24 h after nitrogen stepdown. *hetP*, *hetR*, *hetZ*, *patU5-patU3*, *patA*, *patS,* and most genes of the *hgl* gene cluster were deleted in the septuple mutant.

The structure of σ^{HetZ}-RPo shows that σ^{HetZ} interacts with RNAP core enzyme through the ECF- σ-factor-conserved domains and the σ^{HetZ}-specific insertions. Similar to other ECF σ factors, σ^{HetZ} docks its σ_2 and σ_4 domains on the coiled-coil (CC) and flap-tip helix (FTH) of the RNAP β′1 and β subunit (Fig. 5 *F* and *H*), respectively, and the $\sigma^{HetZ}_{3.2}$ domain threads the active-center cleft of RNAP (Fig. 5*G*). Different from other ECF σ factors, σ^{HetZ} makes additional interactions with RNAP core enzyme, in particular the β domain, using its σ_{i3} domain (Fig. 5*H*).

The σ^{HetZ}_{i2} domain unwinds the promoter DNA at the −12 position (Fig. 6 *A–E* and *SI Appendix*, Fig. S7 *A*, *D*, *H*, and *M*). The cryo-EM map shows that the nucleotides at the −13 position are base paired while the nontemplate nucleotides starting from −12 are unwound, albeit of the complementary sequence at the −12 and −11 positions, indicating that the RNAP- σ^{HetZ} holoenzyme unwinds promoter at the −13/−12 junction. In contrast, other σ^{70}-type factors use the σ_2 domain to unwind promoter DNA (Fig. 6*B* and *SI Appendix*, Fig. S7 *A–C*). A β hairpin of the σ^{HetZ}_{i2} domain protrudes out at the junction and prevents the upstream dsDNA from further extending (Fig. 6*B*). The four hydrophobic residues (I177, L179, L186, L189) on the β hairpin stack on the −13 nucleotide, stabilizing the unwound upstream junction of the transcription bubble (Fig. 6 *B* and *C* and *SI Appendix*, Fig. S7*H*). Deletion of σ^{HetZ}_{i2} severely decreased the transcription activity, supporting its essential role in promoter unwinding (*SI Appendix*, Fig. S7*K*). In summary, σ^{HetZ} uses its σ^{HetZ}_{i2} insertion to unwind promoter DNA, in sharp contrast to other σ^{70}-type factors that use the σ_2 domain for promoter DNA unwinding.

The σ^{HetZ}-recognized promoter has a consensus sequence pattern of "TGnA" at the −10 element (Fig. 1*C*, 546 σ^{HetZ}-recognized promoters). Replacing the T_{-13}, G_{-12}, T_{-11}, or A_{-10} of the *patS* promoter with other bases severely impaired the transcriptional activity, supporting the sequence consensus pattern. However, both T_{-13} and T_{-11} can be replaced with A, and T_{-13} can be partially replaced with G and C, T_{-11} partially replaced with C (Fig. 1*C* and *SI Appendix*, Fig. S7*M*). Our structure reveals that the −10 element is recognized by σ^{HetZ}_{i1} and σ^{HetZ}_{i2} domains. The unwound nucleotides of G_{-12} (nt) and T_{-11}(nt) flip out and insert their bases into a protein pocket formed by R57, I58, E61, E147, and R175 (Fig. 6 *A* and *D* and *SI Appendix*, Fig. S7*I*), where the G_{-12} (nt) and T_{-11}(nt) are recognized (Fig. 6*E* and *SI Appendix*, Fig. S7*I*). The T_{-13}(nt) and A_{-10}(nt) are likely recognized by R190 and R82, respectively, through base-specific H-bond interactions (Fig. 6 *C* and *E* and *SI Appendix*, Fig. S7 *H* and *I*). Alanine substitution of the base-contacting residues substantially impaired the transcription activity, supporting their promoter-recognition function (*SI Appendix*, Fig. S7*K*). It is worth noting that the −10 element is not recognized by the "specificity loop" of σ^{HetZ}, unlike other σ^{70}-type factors (*SI Appendix*, Fig. S7 *D–F*) (60–64).

The σ^{HetZ}_4 domain recognizes the consensus −35 element ($T_{-35}C_{-34}C_{-33}G_{-32}G_{-31}A_{-30}$) (Fig. 6*F*), previously named DIF1 motif (39). Mutation at each position of the consensus sequence resulted in substantial reduction of transcriptional activity, supporting the importance of the −35 element for promoter recognition and transcription initiation, albeit T_{-35}, G_{-32}, and A_{-30} can be fully or partially replaced with alternative bases (Fig. 1*C* and *SI Appendix*, Fig. S7*M*). One of the helices of σ^{HetZ}_4 is poised in the major groove of double-stranded DNA to recognize base moieties of the −35 element (Fig. 6 *F* and *G*). R291, D294, Y295, and Q298 potentially make base-specific interactions with the sequence motif, accounting for the sequence-specific recognition (Fig. 6*G* and *SI Appendix*, Fig. S7*J*). The proposed interaction is supported by the mutation result, showing that alanine

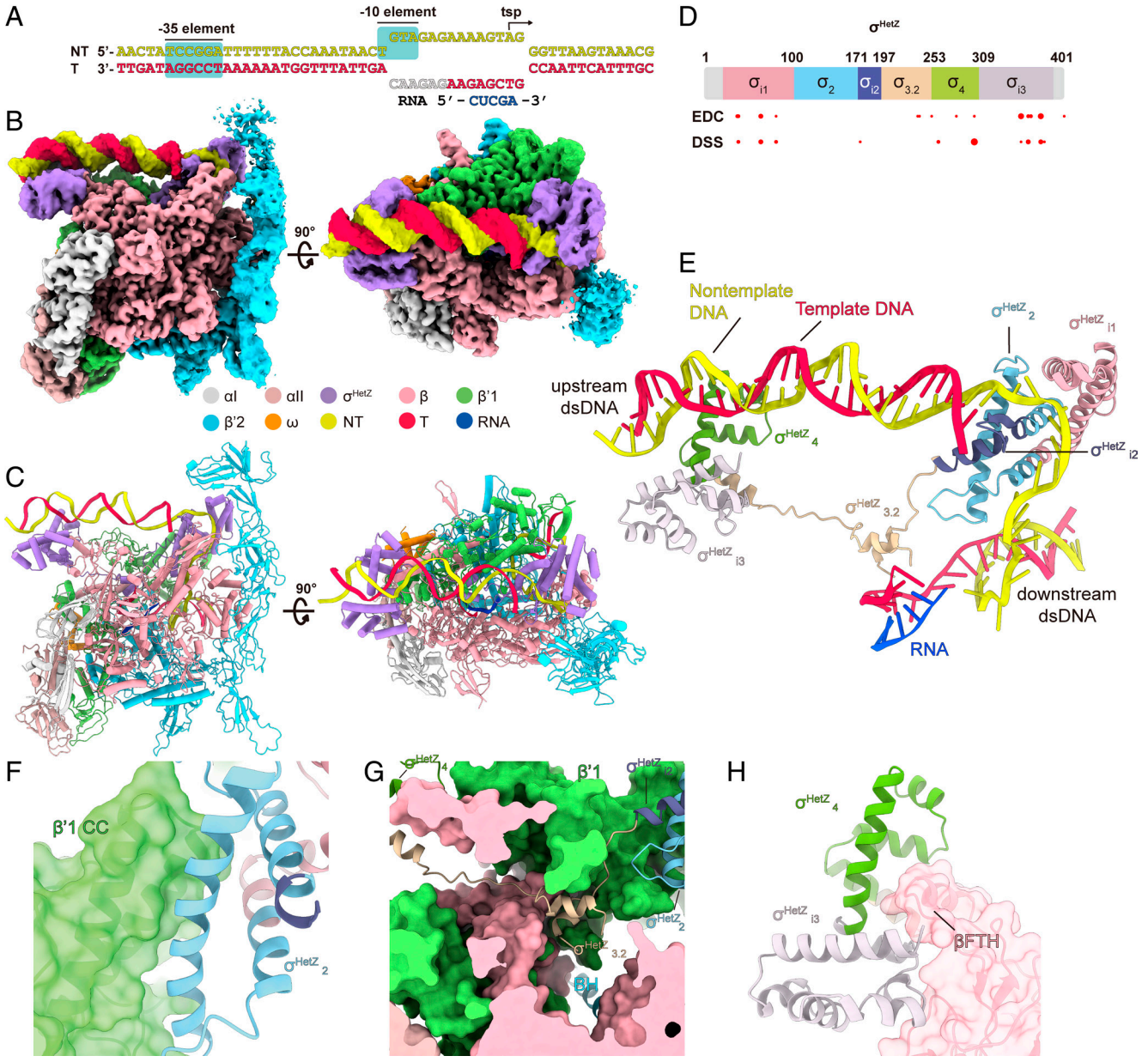

Fig. 5. The overall structure of σ^{HetZ}-Rpo. (*A*) The nucleic-acid scaffold used in the cryo-EM structure determination. (*B*) The front and top views of the cryo-EM map for the structure of σ^{HetZ}-RPo. (*C*) The front and top view orientations of σ^{HetZ}-RPo structure. (*D*) Schematic diagram of σ^{HetZ}, with indications of sites for complexing with PatU3. The sizes of red dots are proportional to times of interactions with PatU3 (Dataset S3). DSS (disuccinimidyl suberate) and EDC (1-ethyl-3-(3-dimethylaminopropyl) are the cross-linkers. (*E*) The interactions between promoter DNA and σ^{HetZ}. The RNAP core enzyme was omitted for clarity. (*F*) The interaction between $\sigma^{HetZ}{}_2$ and the β′1CC domain of RNAP core enzyme. (*G*) The interaction between $\sigma^{HetZ}{}_{3.2}$ and RNAP core enzyme. (*H*) The interaction between $\sigma^{HetZ}{}_4$ and $\sigma^{HetZ}{}_{i3}$ of σ^{HetZ} and RNAP core enzyme.

substitution of R291, D294, and Q298 substantially reduced the transcription activity (*SI Appendix*, Fig. S7*K*). $\sigma^{HetZ}{}_4$ and $\sigma^{HetZ}{}_{i3}$ also make extensive sequence nonspecific polar interactions with the phosphate backbones of the −35 element (*SI Appendix*, Fig. S7*G*).

The activity of σ^{HetZ} is inhibited by the potential anti- σ factor PatU3. Cross-linking mass spectrometry (XL-MS) analyses indicated that PatU3 complexes with σ^{HetZ} through interactions with domains σ_{i1}, σ_{i3} and σ_4 (Fig. 5*D*), preventing σ^{HetZ} from recognizing −10 and −35 regions of target promoters.

Discussion

HetR and PatS-derived RGSGR-containing peptides are the activator and diffusible inhibitor of the long-standing model for heterocyst differentiation and de novo patterning (1, 65, 66). HetZ may activate the expression of *hetR* and *patS* (24), therefore is a strong candidate for a member of the CRC. In this article, we present fourfold evidence to show that HetZ is an ECF sigma factor that directly activates the expression of *patS*, including ChIP-seq, RNA-seq, transcription assay system in *E. coli*, and cryo-EM structure. Similarly, σ^{HetZ} activates *hetR*, *patU5-patU3* (*patU*), *sigC*, *hep* genes, and *ntcA*. Because *hetZ* (σ^{HetZ}) is directly regulated by HetR (23), the "autoregulation" of *hetR* (67), the HetR-dependent activation of *patS* (65), and the mutual regulation between HetR and NtcA (43), are all mediated by σ^{HetZ}, as depicted in Fig. 2*B*. Previously, "DIF1 motif" sequence, typically TCCGGA, has been identified upstream of *hetR*, *sigC*, some *hep* genes (39), and *patS*/*patX* (20), but what interacts with this motif

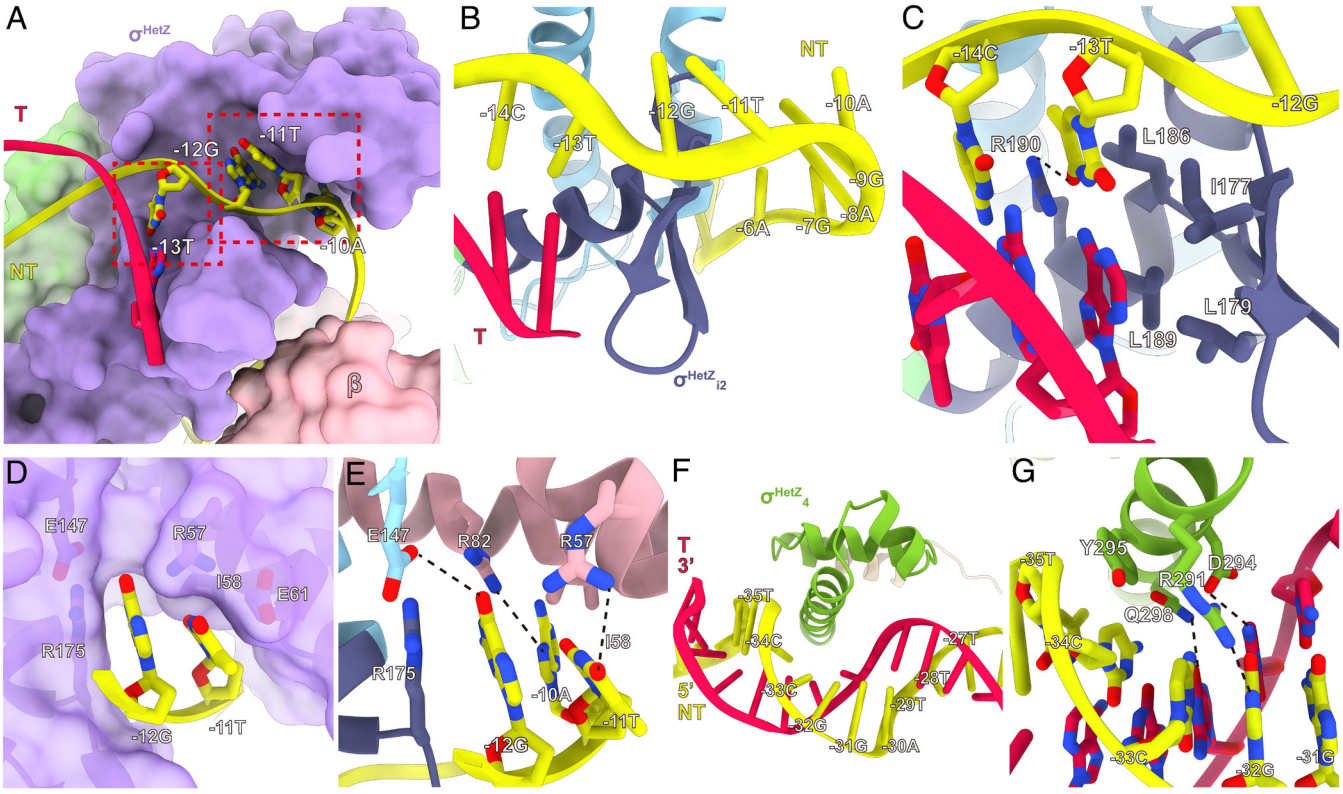

Fig. 6. The interactions between σ^{HetZ} and promoter DNA. (*A*) The interactions between the −10 element and the σ^{HetZ}. Yellow, the nontemplate-strand DNA; red, the template-strand DNA. (*B*) The σ^{HetZ}_{i2} facilitates the promoter unwinding. (*C*) The detailed interactions of σ^{HetZ}_{i2} with the −13 base pair. (*D* and *E*) The detailed interactions between the residues of σ^{HetZ} pocket and the nontemplate nucleotides G_{-12} (nt), T_{-11} (nt), and A_{-10} (nt). (*F* and *G*), The interactions between the −35 element and σ^{HetZ}_4.

or regulates DIF1 promoters was unknown. With the multifold evidence, now we establish that σ^{HetZ} activates the expression from promoters with "DIF1 motif" at the −35 region, including the promoter of *ntcA* with nontypical σ^{HetZ}-recognized sequence. Unambiguously, σ^{HetZ} is an essential factor in the CRC and the central activator in the whole het-GRN (Fig. 2*B*). The leaky phenotypes of some *hetZ* (σ^{HetZ}) mutants is due to the presence of $\sigma^{HetZ'}$ (*SI Appendix*, Fig. S4*B*).

There are two activator–inhibitor pairs, namely HetR-PatS/PatX and σ^{HetZ}-PatU, in the CRC, where the mutual regulation or interactions between these factors/genes form closed loops (Fig. 2*B*). PatU3 in *Anabaena* 7120, or PatU in most other filamentous cyanobacteria, is a potential anti- σ^{HetZ} factor. Through σ^{HetZ}-PatU3 and PatU3-Ftn6 interactions, cell differentiation, and cell division may be coordinated (68). In the Δ*hetP*Δ*hetR*Δ*hetZ*Δ(*patU5-patU*)Δ*patS*Δ*patA*Δ(*hgl* island) septuple deletion mutant of *Anabaena* 7120, we showed that the (HetR⊢PatX)↔(σ^{HetZ}⊢PatU) regulatory circuit was sufficient to initiate cell differentiation, while (HetR⊢PatX/PatS)↔ (σ^{HetZ}⊢PatU) could further resolve the pattern of cell differentiation. In the symbolic expressions, ⊢ stands for inhibition, ↔ for mutual activation. Decoupling the HetR-*hetZ* regulatory relationship disabled the function of the reconstituted minimal CRCs. For the role in enhancement of σ^{HetZ}-activated transcription, HetP is reasonably included in the CRC. NtcA could be either included in the CRC, because the σ^{HetZ}-activated expression of *ntcA* may boost the upregulation of *hetR* through NtcA→(NrrA/Pkn22)→HetR, or excluded from the CRC, because this NtcA→HetR pathway is probably missing in N-nonresponsive species, such as *Cronbergia*.

σ^{HetZ} is a novel type of ECF σ factor with 3 insertion domains, and the unique characteristics of σ^{HetZ} are mainly attributed to σ^{HetZ}_{i1} and σ^{HetZ}_{i2} domains: σ^{HetZ}_{i1} and σ^{HetZ}_{i2} are involved in recognition, σ^{HetZ}_{i2} in unwinding, at the −10 region (Figs. 5 and 6). In other σ factors, these functions are performed by the σ_2 domain (63). Interactions between σ^{HetZ} and PatU3 also involve two of the insertion domains, σ^{HetZ}_{i1} and σ^{HetZ}_{i3}. Therefore, the structure of σ^{HetZ} lays the foundation for the unique mechanism of σ^{HetZ}-activated transcription and the σ^{HetZ}-PatU interaction. Through mutual regulation or interaction with HetR, PatS, and PatU, σ^{HetZ} coordinates cell fate determination; through activation of *hep* genes, *sigC* (then *sigE*) and *ntcA* (then *devH*) (Fig. 2*B*), σ^{HetZ} turns on morphogenesis after cell fate determination.

The reconstructed het-GRN, in particular the CRC and control of downstream gene expression by the CRC (Fig. 2*B*), depicts the general mechanism of cell differentiation for spatial separation of N_2 fixation from oxygenic photosynthesis, providing the molecular basis for understanding the ecological success of this group of multicellular cyanobacteria. On the other hand, heterocyst-forming cyanobacteria may be modified through synthetic biological manipulations and delivered into plant cells, to create vertically transferred nitrogen-fixing symbionts. The CRC could serve as one of the targets for genetic modifications to generate ideal phenotypes, such as short filaments with terminal heterocysts, as in *Richelia* (2).

Materials and Methods

Detailed materials and methods are available in *SI Appendix*.

Growth of Cyanobacterial Strains, Induction of Heterocysts, and Observations of Cellular/Subcellular Structures. Cyanobacterial strains were cultured in BG11 at 30 °C in the light. Antibiotics were added to the medium as appropriate. Heterocyst differentiation was induced in $BG11_0$ ($BG11_0$ without nitrate).

To turn off copper-regulated gene expression from P_{petE}, *Anabaena* cells were washed and resuspended with copper-free $BG11_0$. Microscopy was performed using an Olympus BX41 microscope with a FluoCa CCD color video camera. GFP fluorescence was observed with a Sapphire GFP filter set; OFP fluorescence was observed with a mKO/mOrange bandpass filter set; autofluorescence was observed using the red long pass WG fluorescence cube.

Construction of Plasmids and *Anabaena* Strains. Molecular cloning manipulations were performed following standard procedures or manufacturers' instructions. Construction of plasmids in *Saccharomyces cerevisiae* was performed through cotransformation of BamHI-cut pGF vector and purified PCR fragments. Cyanobacterial DNA was prepared according to a quick miniprep method. Cloned PCR fragments were confirmed by sequencing.

Plasmids were transferred from *E. coli* to *Anabaena* 7120 through conjugation. Mutants of *Anabaena* 7120 were generated through homologous double crossover using a CRISPR/Cpf1 system on a replicative plasmid. Complementation was performed by transfer of a plasmid with the wild-type gene into the mutant. To visualize the promoter activity in different types of cells, plasmids carrying promoter-*gfp* were introduced into *Anabaena* 7120 and the *hetZ*4-201 mutant.

Plasmids and *Anabaena* strains are listed in *SI Appendix*, Table S1.

Western Blot Detection. *Anabaena* cells were collected, washed with TE supplemented with Protease Inhibitor Cocktail. Cells were broken with a French Press and the cell debris and membranes were removed by centrifugation. Proteins were separated by SDS-PAGE and electroblotted onto nitrocellulose membranes. σ^{HetZ} was detected with purified rabbit anti- σ^{HetZ} IgG antibody, using the horse radish peroxidase conjugated secondary antibody with luminol as the substrate.

ChIP-seq Analyses. After 8 h of nitrogen stepdown, *Anabaena* cells were cross-linked with 1% formaldehyde for 15 min, and the reaction was terminated by addition of glycine. Cells suspended in cold D-PBS buffer with protease inhibitors were broken by passing through a French press. The cell lysates were then subjected to sonication to shear the DNA into 200 to 500 bp fragments. The sheared DNA was diluted with ChIP dilution buffer, mixed with Dynabeads™ Protein G preloaded with antibody and incubated at 4 °C. Beads with antibody–protein–DNA complexes were washed with washing buffer. Bound material was then eluted from the beads, treated with RNase A, followed by proteinase K. Immunoprecipitated DNA was used to construct sequencing libraries for Illumina® sequencing.

Clean reads were mapped to the *Anabaena* 7120 genome. MACS2 software version 2.1.1 was used to call peaks. Differential peak analyses were conducted using DiffBind package, focusing on the peaks identified in the wild type. The analyses were based on 3 biological replicates. Genome-wide mapping of binding profiles was visualized using Circos (v0.69-9), and visualization of peaks was performed using Gviz.

RNA-seq Analyses. Total RNA was extracted using the TRIzol® Reagent and treated with DNase I to remove the genomic DNA. Strand-specific RNA-seq libraries were prepared with total RNA using MGIEasy RNA Directional Library Prep Set after removing rRNA. Libraries were sequenced on MGISEQ-2000.

Clean reads were aligned to reference genome using TopHat2. Raw count data and FPKM values were generated from the alignment files using bam2rpkm. Raw count data were then used as input into DESeq2 for differential expression analyses (3 biological replicates). Genes with a fold change ≥ 2 or ≤ 0.5 and q-value < 0.01 were taken as differentially expressed between samples.

Protein Purification, Cryo-EM Data Collection, and Processing. The recombinant *Anabaena* RNAP core enzyme was purified from *E. coli* BL21(DE3) cells carrying pCDFDuet-1/*rpoA-rpoZ* and pETDuet/*rpoB-rpoC1-rpoC2* using Ni-NTA affinity and ion-exchange chromatography. The recombinant *Anabaena* σ^{HetZ} was purified from *E. coli* BL21(DE3) carrying pET21b-*hetZ* using a Ni-NTA agarose column. The nucleic-acid scaffold used for the cryo-EM structure of σ^{HetZ}-RPo contains a template DNA (5′-AACTATCCGGATTTTTTACCAAATAACTGTAGAGAAAAGTAGGGTTAAGTAAACG-3′), a nontemplate DNA (5′-TTGATAGGCCTAAAAAATGGTTTATTGACAAGAGAAGAGCTGCCAATTCATTTGC-3′), and an RNA (5′-CUCGA-3′).

Cryo-EM data were collected using EPU in the superresolution mode on a 300 kV Titan Krios (FEI) equipped with a Gatan K3 Summit direct electron detector. A total of 3,216 images were recorded with a pixel size of 0.89 Å and a dose rate of 19 electrons/pixel/s. 3,478,184 particles were picked from motion-corrected images with Blob Picker and extracted with a box size of 150 px with Particle Extraction, subjected to 2D (N = 100) and 3D classification. The best 3D class containing a total of 192,944 particles was polished with Bayesian Polishing in RELION 3.1, refined with Homo refinement and nonuniform refinement in cryoSPARC (v4.2.1), resulting in a nominal resolution of 2.88 Å. The statistics of structure determination are summarized in *SI Appendix*, Table S2.

Radiochemical In Vitro Transcription Assays. The transcription activity of *Anabaena* σ^{HetZ} was measured in reaction mixtures containing the RNAP core enzyme, σ^{A} or σ^{HetZ}, PCR-generated $P_{patS'}$ and the NTP mixture with [α-^{32}P]UTP. The transcripts were then subjected to urea polyacrylamide gel electrophoresis, and the radiograph was obtained by storage-phosphor imaging.

Fluorescence-Detected In Vitro Transcription Assay. The reaction mixtures contained the RNAP core enzyme, *patS* promoter DNA (5′-TAGTAAGCCTTTTACTTCTCCCCGTTAGCTTTTTGCATTTAACTATCCGGATTTTTTACCAAATAACTGTAGAGAAAAGTAGGGTTAAGTAAACGGGCAAATTTTCGTGAAAACAATGTTTTATACATATGGCACGTACGAAGGAAGGATTGGTATGTGGTATATTCGTACGTGCCGGCCTGCTGGTAATCGCAGGCCTTTTTATTT-3′) and σ^{HetZ} or its derivatives. The NTP mix and TOI-3PEG-Biotin were added to the mixture to initiate the reactions. The fluorescence signals were measured using a Tecan Spark plate reader with excitation at 510 nm and emission at 550 nm. The results are presented as mean ± SD (3 biological replicates).

XL-MS of σ^{HetZ}-PatU3 Complex. Ni-NTA affinity chromatography was employed to purify the σ^{HetZ}-PatU3 complex from *E. coli* Rosetta (DE3) expressing the σ^{HetZ} (aa 12-401)-$(Gly\text{-}Ser)_6$-PatU3 (aa 2-258) fusion protein. The eluted protein was treated with TEV protease to remove the N-terminal His_8-SUMO tag. For XL-MS, the target protein was further purified using a Superdex 200 Increase 10/300 GL size-exclusion column.

Cross-linking reactions within the σ^{HetZ}-PatU3 complex were carried out with disuccinimidyl suberate (DSS) or 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC). The reaction was then quenched with NH_4HCO_3. Proteins were precipitated with ice-cold acetone, alkylated with iodoacetamide, and digested with trypsin. LC-MS/MS analyses were performed on an Easy-nLC 1000 II HPLC coupled to a Q-Exactive HF mass spectrometer. Each sample was analyzed twice, and the two technical repeats were combined for data analysis. pLink 2 software was utilized to identify cross-linked peptides, and the results were filtered by applying a 5% FDR cutoff at the spectral level and then further filtered with E-value cutoff < 0.001 and spectral count ≥ 3.

Quantitation of GFP Expression in *E. coli*. *E. coli* DH10B harboring pHB7508 or pHB8140, and the plasmid with promoter-*gfp*, with or without pACYC184 expressing a σ factor gene (Fig. 2*A*), was streaked on LB plates and grown in liquid LB medium at 37 °C overnight. The *E. coli* cells were inoculated into LB medium; appropriate antibiotics, 1 mM L-arabinose, and/or 1 mM IPTG were added to the medium as needed. GFP fluorescence images were taken on a Leica TCS SP8 confocal microscope, using a 405 nm laser line for excitation and a HyD detector for collection of 500 to 560 nm emission. Using the Leica Application Suite X 3.1.1.15751 software, fluorescence intensities were measured from sampling areas within cells. The relative intensity per pixel values of 11 to 21 cells for each strain were used for evaluation of GFP expression and statistical analyses.

Identification of the Conserved Sequence of σ^{HetZ}-Recognized Promoters. Sequences of 9 σ^{HetZ}-recognized promoters were aligned using T-Coffee (v11.00), generating the consensus sequence $TCCGGAN_{16\text{-}17}TGNAN_{7\text{-}9}T$ (N represents A, T, G, or C) (*SI Appendix*, Fig. S1*D*). Further, genomic regions of *hetZ* mutant vs. WT ChIP-seq differential peaks were scanned for the consensus sequence, allowing no more than two mismatches. A total of 546 peaks with σ^{HetZ}-recognized promoters were identified (Dataset S2). These promoter sequences were aligned using GLAM2 in MEME Suite 5.5.7 to generate the consensus sequence.

Construction of the Phylogenetic Tree of σ^{HetZ} and Homologs. Protein sequences of σ^{HetZ} and homologs were aligned using MUSCLE, and the phylogenetic analyses were performed in MEGA7 using the Maximum Likelihood method.

Data, Materials, and Software Availability. RNA-seq raw reads for the wild type and mutants of *Anabaena* 7120 have been deposited in the NCBI SRA under accession numbers SRR32280515, SRR32280516, SRR32280518, SRR32280520-SRR32280522, and SRR32993896-SRR32993898 (69). σ^{HetZ} ChIP-seq raw reads for *Anabaena* 7120 strains have been deposited in the NCBI SRA under accession numbers SRR32409265-SRR32409280, SRR32409283,

and SRR32409284 (70). ChIP-seq processed data have been deposited in the NCBI GEO under accession number GSE310072 (71). All other data are included in the manuscript and/or supporting information.

ACKNOWLEDGMENTS. This work was supported by the National Key Research and Development Program of China (2021YFA0909700, 2021YFA09 10700) and the National Natural Science Foundation of China (32170054, 31770044). We thank L. Kong, Y. Song, G. Li, J. Duan at the electron microscopy system of the National Facility for Protein Science in Shanghai for technical support and assistance in cryo-EM data Collection, Z. Zhang and M. Zhang at the core facility of the Center for Excellence in Molecular Plant Sciences for help with cryo-electron microscopy (EM) sample screening, Y. Cao at Shanghai Institute of Immunity and Infection for assistance with the XL-MS experiment.

1. C. P. Wolk, A. Ernst, J. Elhai, "Heterocyst metabolism and development" in *The Molecular Biology of Cyanobacteria*, D. A. Bryant, Ed. (Springer Netherlands, Dordrecht, 1994), pp. 769–823.
2. J. A. Hilton *et al.*, Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. *Nat. Commun.* **4**, 1767 (2013).
3. C. Álvarez *et al.*, Symbiosis between cyanobacteria and plants: From molecular studies to agronomic applications. *J. Exp. Bot.* **74**, 6145–6157 (2023).
4. H. Meinhardt, A. Gierer, Pattern formation by local self-activation and lateral inhibition. *Bioessays* **22**, 753–760 (2000).
5. B. J. MacGregor *et al.*, Microbiological, molecular biological and stable isotopic evidence for nitrogen fixation in the open waters of Lake Michigan. *Environ. Microbiol.* **3**, 205–219 (2001).
6. J. P. Zehr, Nitrogen fixation by marine cyanobacteria. *Trends Microbiol.* **19**, 162–173 (2011).
7. K. Tang *et al.*, Spatial distribution and core community of diazotrophs in biological soil crusts and subsoils in temperate semi-arid and arid deserts of China. *Front. Microbiol.* **14**, 1074855 (2023).
8. C. P. Wolk *et al.*, Isolation and complementation of mutants of *Anabaena* sp. strain PCC 7120 unable to grow aerobically on dinitrogen. *J. Bacteriol.* **170**, 1239–1244 (1988).
9. M. A. Murry, C. P. Wolk, Evidence that the barrier to the penetration of oxygen into heterocysts depends upon two layers of the cell envelope. *Arch. Microbiol.* **151**, 469–474 (1989).
10. A. Magnuson, T. Cardona, Thylakoid membrane function in heterocysts. *Biochim. Biophys. Acta* **1857**, 309–319 (2016).
11. A. Valladares *et al.*, Cytochrome *c* oxidase genes required for nitrogenase activity and diazotrophic growth in *Anabaena* sp. PCC 7120. *Mol. Microbiol.* **47**, 1239–1249 (2003).
12. M. Ermakova *et al.*, Heterocyst-specific flavodiiron protein Flv3B enables oxic diazotrophic growth of the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11205–11210 (2014).
13. S. Laurent *et al.*, Nonmetabolizable analogue of 2-oxoglutarate elicits heterocyst differentiation under repressive conditions in *Anabaena* sp. PCC 7120. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9907–9912 (2005).
14. M.-X. Zhao *et al.*, Structural basis for the allosteric control of the global transcription factor NtcA by the nitrogen starvation signal 2-oxoglutarate. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12487–12492 (2010).
15. W. J. Buikema, R. Haselkorn, Expression of the *Anabaena hetR* gene from a copper-regulated promoter leads to heterocyst differentiation under repressing conditions. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2729–2734 (2001).
16. Y. Kim *et al.*, Structures of complexes comprised of *Fischerella* transcription factor HetR with *Anabaena* DNA targets. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1716–E1723 (2013).
17. J. Komarek, E. Zapomělová, F. Hindak, *Cronbergia* gen. nov., a new cyanobacterial genus (Cyanophyta) with a special strategy of heterocyte formation. *Cryptogamie Algol.* **31**, 321–341 (2010).
18. L. Corrales-Guerrero, V. Mariscal, E. Flores, A. Herrero, Functional dissection and evidence for intercellular transfer of the heterocyst-differentiation PatS morphogen. *Mol. Microbiol.* **88**, 1093–1105 (2013).
19. H.-X. Hu *et al.*, Structural insights into HetR—PatS interaction involved in cyanobacterial pattern formation. *Sci. Rep* **5**, 16470 (2015).
20. J. Elhai, I. Khudyakov, Ancient association of cyanobacterial multicellularity with the regulator HetR and an RGSGR pentapeptide-containing protein (PatX). *Mol. Microbiol.* **110**, 931–954 (2018).
21. H.-S. Yoon, J. W. Golden, Heterocyst pattern formation controlled by a diffusible peptide. *Science* **282**, 935–938 (1998).
22. I. Khudyakov, G. Gladkov, J. Elhai, Inactivation of three RG(S/T)GR pentapeptide-containing negative regulators of HetR results in lethal differentiation of *Anabaena* PCC 7120. *Life* **10**, 326 (2020).
23. Y. Du, Y. Cai, S. Hou, X. Xu, Identification of the HetR recognition sequence upstream of *hetZ* in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **194**, 2297–2306 (2012).
24. Y. Du *et al.*, Expression from DIF1-motif promoters of *hetR* and *patS* is dependent on HetZ and modulated by PatU3 during heterocyst differentiation. *PLoS One* **15**, e0232383 (2020).
25. P. Videau *et al.*, The *hetZ* gene indirectly regulates heterocyst development at the level of pattern formation in *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **109**, 91–104 (2018).
26. X. Zeng, C. C. Zhang, The making of a heterocyst in cyanobacteria. *Annu. Rev. Microbiol.* **76**, 597–618 (2022).
27. S. Deora, G. S. Deora, S. Nigam, Harish, Hacking heterocysts: Advances in the genetic regulation of heterocyst differentiation. *Arch. Microbiol.* **208**, 80 (2025).
28. G. Huang *et al.*, Clustered genes required for the synthesis of heterocyst envelope polysaccharide in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **187**, 1114–1123 (2005).
29. Y. Wang *et al.*, Predicted glycosyl transferase genes located outside the HEP island are required for formation of heterocyst envelope polysaccharide in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **189**, 5372–5378 (2007).
30. G. Fiedler, M. Arnold, S. Hannus, I. Maldener, The DevBCA exporter is essential for envelope formation in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **27**, 1193–1202 (1998).
31. Q. Fan *et al.*, Clustered genes required for synthesis and deposition of envelope glycolipids in *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **58**, 227–243 (2005).
32. D. Shvarev, C. N. Nishi, L. Wörmer, I. Maldener, The ABC transporter components HgdB and HgdC are important for glycolipid layer composition and function of heterocysts in *Anabaena* sp. PCC 7120. *Life* **8**, 26 (2018).
33. K. C. Higa *et al.*, The RGSGR amino acid motif of the intercellular signalling protein, HetN, is required for patterning of heterocysts in *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **83**, 682–693 (2012).
34. T. Thiel, Organization and regulation of cyanobacterial nif gene clusters: Implications for nitrogenase expression in plant cells. *FEMS Microbiol. Lett.* **366**, fnz077 (2019).
35. A. Valladares, E. Flores, A. Herrero, Transcription activation by NtcA and 2-oxoglutarate of three genes involved in heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **190**, 6126–6133 (2008).
36. X. Xu *et al.*, Functional dissection of the CRP-family transcription factor DevH and its interplay with NtcA in a cyanobacterium. *Cell Rep.* **44**, 116435 (2025).
37. W. Zhang *et al.*, A gene cluster that regulates both heterocyst differentiation and pattern formation in *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **66**, 1429–1443 (2007).
38. R. Rachedi *et al.*, Unravelling HetC as a peptidase-based ABC exporter driving functional cell differentiation in the cyanobacterium *Nostoc* PCC 7120. *Microbiol. Spectr.* **12**, e0405823 (2024).
39. J. Mitschke *et al.*, Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20130–20135 (2011).
40. J. Guío *et al.*, 2-oxoglutarate modulates the affinity of FurA for the *ntcA* promoter in *Anabaena* sp. PCC 7120. *FEBS Lett.* **594**, 278–289 (2020).
41. J. D. Helmann, The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol.* **46**, 47–110 (2002).
42. T. Mascher, Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. *Curr. Opin. Microbiol.* **16**, 148–155 (2013).
43. A. M. Muro-Pastor, A. Valladares, E. Flores, A. Herrero, Mutual dependence of the expression of the cell differentiation regulatory protein HetR and the global nitrogen regulator NtcA during heterocyst development. *Mol. Microbiol.* **44**, 1377–1385 (2002).
44. H. Zhang, S. Wang, Y. Wang, X. Xu, Functional overlap of *hetP* and *hetZ* in regulation of heterocyst differentiation in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **200**, e00707-17 (2018).
45. S. Picossi, E. Flores, A. Herrero, ChIP analysis unravels an exceptionally wide distribution of DNA binding sites for the NtcA transcription factor in a heterocyst-forming cyanobacterium. *BMC Genomics* **15**, 22 (2014).
46. R. A. Mella-Herrera *et al.*, The *sigE* gene is required for normal expression of heterocyst-specific genes in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **193**, 1823–1832 (2011).
47. S. Ehira, M. Ohmori, NrrA, a nitrogen-responsive response regulator facilitates heterocyst development in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **59**, 1692–1703 (2006).
48. S. Ehira, M. Ohmori, NrrA directly regulates expression of *hetR* during heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **188**, 8520–8525 (2006).
49. B. Roumezi *et al.*, The Pkn22 kinase of *Nostoc* PCC 7120 is required for cell differentiation via the phosphorylation of HetR on a residue highly conserved in genomes of heterocyst-forming cyanobacteria. *Front Microbiol* **10**, 3140 (2019).
50. Y. Kurio *et al.*, The CRP-family transcriptional regulator DevH regulates expression of heterocyst-specific genes at the later stage of differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **114**, 553–562 (2020).
51. S. Hou *et al.*, The HetR-binding site that activates expression of *patA* in vegetative cells is required for normal heterocyst patterning in *Anabaena* sp. PCC 7120. *Sci. Bull.* **60**, 192–201 (2015).
52. I. Y. Khudyakov, J. W. Golden, Identification and inactivation of three group 2 sigma factor genes in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **183**, 6667–6675 (2001).
53. M. R. Aldea, R. A. Mella-Herrera, J. W. Golden, Sigma factor genes *sigC*, *sigE*, and *sigG* are upregulated in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **189**, 8392–8396 (2007).
54. S. Ehira, S. Miyazaki, Regulation of genes involved in heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120 by a group 2 sigma factor SigC. *Life* **5**, 587–603 (2015).
55. J. Liang, L. Scappino, R. Haselkorn, The *patA* gene product, which contains a region similar to CheY of *Escherichia coli*, controls heterocyst pattern formation in the cyanobacterium *Anabaena* 7120. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5655–5659 (1992).
56. K. C. Higa, S. M. Callahan, Ectopic expression of *hetP* can partially bypass the need for *hetR* in heterocyst differentiation by *Anabaena* sp. strain PCC 7120. *Mol. Microbiol.* **77**, 562–574 (2010).
57. Y.-M. Chang *et al.*, Structural study of TcaR and its complexes with multiple antibiotics from *Staphylococcus epidermidis*. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8617–8622 (2010).
58. Y.-M. Chang *et al.*, Functional studies of ssDNA binding ability of MarR family protein TcaR from *Staphylococcus epidermidis*. *PLoS One* **7**, e45665 (2012).
59. Y.-M. Chang *et al.*, TcaR–ssDNA complex crystal structure reveals new DNA binding mechanism of the MarR family proteins. *Nucleic Acids Res.* **42**, 5314–5321 (2014).
60. A. Feklistov, S. A. Darst, Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* **147**, 1257–1269 (2011).
61. Y. Zhang *et al.*, Structural basis of transcription initiation. *Science* **338**, 1076–1080 (2012).
62. S. Campagne *et al.*, Structural basis for−10 promoter element melting by environmentally induced sigma factors. *Nat. Struct. Mol. Biol.* **21**, 269–276 (2014).
63. J. Chen, H. Boyaci, E. A. Campbell, Diverse and unified mechanisms of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **19**, 95–109 (2021).
64. L. Li *et al.*, Structural basis for transcription initiation by bacterial ECF σ factors. *Nat. Commun.* **10**, 1153 (2019).
65. X. Huang, Y. Dong, J. Zhao, HetR homodimer is a DNA-binding protein required for heterocyst differentiation, and the DNA-binding activity is inhibited by PatS. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4848–4853 (2004).

66. D. D. Risser, S. M. Callahan, Genetic and cytological evidence that heterocyst patterning is regulated by inhibitor gradients that promote activator decay. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19884–19888 (2009).
67. T. A. Black, Y. Cai, C. P. Wolk, Spatial expression and autoregulation of *hetR*, a gene involved in the control of heterocyst development in *Anabaena*. *Mol. Microbiol.* **9**, 77–84 (1993).
68. L. Yin *et al.*, PatU3 plays a central role in coordinating cell division and differentiation in pattern formation of filamentous cyanobacterium *Nostoc* sp. PCC 7120. *Sci. China Life Sci* **66**, 2896–2909 (2023).
69. Y. Wang, X. Hu, X. Xu, Data from "Nostoc sp. PCC 7120 RNA-seq" NCBI SRA. https://www.ncbi.nlm.nih.gov/sra/?term=SRP562406. Deposited 8 February 2025.
70. Y. Wang, X. Hu, X. Xu, Data from "Nostoc sp. PCC 7120 ChIP-seq" NCBI SRA. https://www.ncbi.nlm.nih.gov/sra/?term=SRP565168. Deposited 20 February 2025.
71. Y. Wang, X. Hu, X. Xu, Integrated sudden origin of complex N2-fixing cells upon global oxygenation. GEO. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE310072. Deposited 20 November 2025.